

Prov. 9. 14/52

T
15
7

UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS



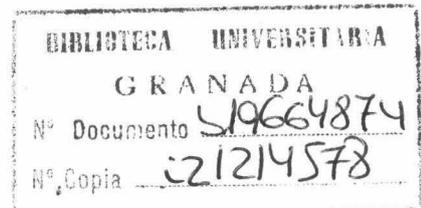
Departamento de Estadística e Investigación Operativa

UN ESTUDIO SOBRE LOS PROBLEMAS DE
ESTIMACIÓN, ALISAMIENTO Y PREDICCIÓN
DE CAMPOS ALEATORIOS

TESIS DOCTORAL

Francisco Javier Alonso Morales.

GRANADA, Mayo de 1994



UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Fecha ... 12-5-94
ENTRADA NUM. ... 727

UN ESTUDIO SOBRE LOS PROBLEMAS DE
ESTIMACIÓN, ALISAMIENTO Y PREDICCIÓN
DE CAMPOS ALEATORIOS

Memoria presentada para optar
al grado de Doctor en Ciencias
(Matemáticas) por
Francisco Javier Alonso Morales.



Vº Bº
Director de Tesis:



Prof. Dr. D. José Miguel Angulo Ibáñez.

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

FACULTAD DE CIENCIAS

UNIVERSIDAD DE GRANADA

UNIVERSIDAD DE GRANADA
- 4 MAYO 1994
COMISION DE DOCTORADO

A mi familia

INTRODUCCIÓN

En muy diversos campos científicos (ciencias medioambientales, físicas, geológicas, procesamiento de imágenes, etc.), se suelen presentar situaciones en las que se dispone de observaciones en espacios de dimensión mayor que uno (por ejemplo, ver Whittle (1954), Besag(1974), Martin (1979), Besag (1986), Unny (1989), Cressie (1991)) e interesa su análisis, especialmente en relación con la estructura espacial de los fenómenos subyacentes. Este hecho ha motivado un creciente interés en el estudio de procesos en varias dimensiones. Sin embargo, la extensión de los resultados disponibles para una a dos o más dimensiones no es, en absoluto, inmediata (Ripley (1981), Guyon (1990)), siendo necesario desarrollar en muchos aspectos una metodología que, en la medida de lo posible, generalice los resultados de que se dispone en una dimensión.

En tal sentido, se han desarrollado en los últimos años modelos y metodologías que, en muchas ocasiones parecen constituir extensiones “naturales” de los relativos a una dimensión: Modelos Markovianos (ver, por ejemplo, Whittle (1954), Besag (1974), Bartlett (1975)), modelos ARMA (Huang (1984), Guyon (1990), Martin (1990), Huang et al. (1992)), procesos definidos por ecuaciones en derivadas parciales estocásticas (Whittle (1954), Cairolì et al. (1975), Nualart et al. (1979), Unny (1989)), etc.. Algunas técnicas de estimación son extensión de las disponibles, por ejemplo, para el análisis de series temporales (Besag (1974), Tjøstheim (1978), Martin (1990), Huang (1992)); otras han sido desarrolladas especialmente para campos aleatorios (aunque una restricción a una dimensión suele ser sencilla) (por ejemplo, los métodos *coding* de Besag (1974)). Además, aparecen conceptos nuevos, como el de separabilidad (Martin (1979)), que simplifica el problema multidimensional, anisotropía, etc., y que, obviamente, no tienen equivalente en una dimensión.

Entre problemas fundamentales que se presentan, se pueden citar la elección de un modelo, la estimación de sus parámetros, pretendiendo con ello el alisamiento de las observaciones y la estimación en puntos donde no se disponga de observación, diseño de redes de observación, etc.. El principal problema que suelen presentar los datos espaciales es que, generalmente, las localizaciones de observación están distribuidas irregularmente. Otras veces éstas se encuentran situadas sobre una malla regular, pero hay un gran número de nodos sin observación real. Adicionalmente, la observación suele llevar un cierto error inherente. Por esto, se deben de procurar métodos que per-

mitan la estimación de los parámetros y problemas derivados cuando no se disponga de una información exhaustiva de todos los nodos. En tal sentido, los algoritmos EM, ECM y MCECM (Dempster et al. (1977), Meng et al. (1993), vanDik et al. (1994)) y algoritmos suplementados, que también calculan las matrices de varianzas-covarianzas asintóticas entre los estimadores, constituyen una metodología muy útil que ha recibido gran atención en los últimos años.

Los objetivos de esta memoria son desarrollar una metodología basada en la filosofía EM para la inferencia sobre campos aleatorios, y en particular, la estimación de los parámetros y el alisamiento de los valores del campo, así como la predicción en los nodos donde no se disponga de observaciones, de acuerdo con la estructura del modelo. Esta metodología se ha inspirado en la desarrollada por Shumway y Stoffer (1982) para series temporales.

En el capítulo uno se describen los fundamentos del algoritmo EM (Dempster et al. (1977)), resultados principales sobre convergencia del algoritmo (Wu (1983), Boyles (1983)) y algunas variantes recientes como los algoritmos ECM, MCECM, SEM, SECM y SMCECM (Meng et al. (1993), Meng (1994), vanDik et al. (1994)).

El capítulo dos se dedica al estudio de modelos espaciales con parámetro discreto, especialmente los modelos ARMA sobre \mathbb{Z}^2 (Besag (1974), Tjøstheim (1978), Martin (1990), Guyon (1990), Cressie (1991), etc.) y algunos modelos más generales.

En el capítulo tres se exponen, desde el punto de vista teórico, los resultados obtenidos en el desarrollo de la metodología para la estimación, alisamiento y predicción de algunos modelos de campos aleatorios, en los que se hace uso de contenidos de los capítulos anteriores.

Ilustramos estos métodos, en el capítulo cuatro, mediante una serie de simulaciones sobre las que aplicamos la metodología descrita en el capítulo tres, poniendo de manifiesto su buen comportamiento incluso en casos en que la información proporcionada por los datos es muy pobre (pocos datos en relación al total). Con tal fin se ha implementado una serie de programas en FORTRAN 77 que hacen uso de rutinas de la librería IMSL.

Al final de la memoria (capítulo cinco), se mencionan algunos enfoques alternativos y abiertos que son objeto de nuestra investigación actual en relación con el contenido de esta memoria.

Antes de acabar, quiero agradecer a todos mis compañeros y amigos del departamento de Estadística e I. O. de esta universidad la colaboración que he

recibido de ellos durante la realización de esta memoria. En particular, quiero mostrar mi agradecimiento a María del Carmen Bueso Sánchez por sus ideas y comentarios sobre algunos aspectos de la misma.

Por último, y muy especialmente, quiero reconocer y agradecer la inestimable ayuda que el director de esta memoria, D. José Miguel Angulo Ibáñez, me ha proporcionado durante toda la realización de la misma, no sólo en el aspecto científico sino también en el aspecto humano como amigo.

Indice

1	Algoritmos EM	1
1.1	Algoritmo EM	1
1.1.1	Introducción	1
1.1.2	Definición del Algoritmo EM	4
1.1.3	Resultados de interés	5
1.2	Alisamiento y predicción de series temporales usando el algoritmo EM	16
1.2.1	Introducción	16
1.2.2	Estimación de los parámetros usando el algoritmo EM	18
1.3	El algoritmo ECM	22
1.4	Algoritmos suplementados	26
2	Elementos sobre campos espaciales con parámetro discreto	33
2.1	Introducción	33
2.2	Procesos ARMA sobre \mathbb{Z}^2	34
2.3	Modelos Markovianos	38
2.3.1	Algunos esquemas espaciales	42
2.4	Procesos separables	43
2.5	Métodos de estimación	45
2.5.1	Estimación máximo-verosímil	45
2.5.2	Métodos coding	46
2.5.3	Estimación de procesos ARMA	48
3	Estimación en cierto tipo de procesos espaciales con datos faltantes	51
3.1	Introducción	51
3.2	Modelo	52

3.3	Adaptación del algoritmo EM	55
3.4	Estimación y alisamiento	61
3.4.1	Problema de identificación	63
3.4.2	Error de predicción	64
3.5	El algoritmo ECM	65
3.6	Modelo $AR(2) \times AR(1)$	67
3.7	Modelo $AR(2) \times AR(2)$	78
3.8	Algunas consideraciones sobre otros modelos	82
3.8.1	Modelo $AR(1) \times MA(1)$	82
3.8.2	Modelo $MA(1) \times MA(1)$	83
3.8.3	Caso general no separable	84
4	Aplicaciones mediante simulación	85
4.1	Introducción	85
4.2	Simulación de los datos	86
4.3	Estimación de los parámetros y alisamiento	87
4.4	Algunas simulaciones estudiadas	88
5	Enfoques alternativos y problemas abiertos	95
5.1	Algunos enfoques alternativos	95
5.2	Otros problemas abiertos	101
A	Datos simulados	103
B	Gráficos	133
C	Organigramas	153

Capítulo 1

Algoritmos EM

1.1 Algoritmo EM

1.1.1 Introducción

El término *datos incompletos*, en su forma general, hace referencia a dos espacios muestrales \mathbf{X} e \mathbf{Y} y una aplicación de \mathbf{X} a \mathbf{Y} . En vez de observar los *datos completos* \mathbf{x} en \mathbf{X} , observamos los *datos incompletos* $\mathbf{y}=\mathbf{y}(\mathbf{x})$ en \mathbf{Y} . En el desarrollo del modelo que proponemos en el capítulo tres, para campos espaciales, \mathbf{x} serán tanto los datos que pretendemos alisar como los datos que se han observado, e \mathbf{y} serán solamente los datos observados. Sea $f(\mathbf{x} | \theta)$ la función de densidad de \mathbf{x} con parámetro $\theta \in \Theta$; la función de densidad de \mathbf{y} vendrá dada por

$$g(\mathbf{y} | \theta) = \int_{X(\mathbf{y})} f(\mathbf{x} | \theta) d\mathbf{x}, \quad (1.1)$$

donde $X(\mathbf{y}) = \{\mathbf{x} : \mathbf{y}(\mathbf{x}) = \mathbf{y}\}$.

El algoritmo EM está dirigido a encontrar un valor de θ que maximice la función anteriormente dada, $g(\mathbf{y} | \theta)$, dado que \mathbf{y} es observado. Esto se lleva a cabo haciendo uso de la familia asociada $f(\mathbf{x} | \theta)$. Observemos que dada la función de densidad de los *datos incompletos* $g(\mathbf{y} | \theta)$ hay varias especificaciones posibles de los *datos completos* $f(\mathbf{x} | \theta)$. A veces será obvia una elección natural, pero otras veces pueden existir diferentes formas de definir $f(\mathbf{x} | \theta)$.

Cada iteración del algoritmo EM involucra dos pasos que se denominarán paso E (paso “*expectation*”, toma de esperanzas condicionadas) y paso M (paso

“*maximization*”, maximización de la función resultante). El siguiente ejemplo numérico, usado por Dempster et al. (1977) para introducir el algoritmo EM, muestra de forma esquemática la aplicación del método.

Se consideran datos de la distribución de 197 animales multinomialmente en cinco categorías $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, donde las probabilidades de cada categoría son

$$\left(\frac{1}{2}, \frac{\pi}{4}, \frac{1-\pi}{4}, \frac{1-\pi}{4}, \frac{\pi}{4}\right),$$

para algún π con $0 \leq \pi \leq 1$.

Supongamos que los *datos observados* son de la forma

$$\mathbf{y} = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34),$$

donde $y_1 = x_1 + x_2$, $y_2 = x_3$, $y_3 = x_4$, $y_4 = x_5$. Es decir, esta variable \mathbf{y} tiene distribución multinomial con probabilidades

$$\left(\frac{1}{2} + \frac{\pi}{4}, \frac{1-\pi}{4}, \frac{1-\pi}{4}, \frac{\pi}{4}\right).$$

Entonces,

$$g(\mathbf{y} | \pi) = \frac{(y_1 + y_2 + y_3 + y_4)!}{y_1!y_2!y_3!y_4!} \left(\frac{1}{2} + \frac{\pi}{4}\right)^{y_1} \left(\frac{1-\pi}{4}\right)^{y_2} \left(\frac{1-\pi}{4}\right)^{y_3} \left(\frac{\pi}{4}\right)^{y_4}.$$

La especificación de la función masa de probabilidad de los *datos completos* es

$$f(\mathbf{x} | \pi) = \frac{(x_1 + x_2 + x_3 + x_4 + x_5)!}{x_1!x_2!x_3!x_4!x_5!} \left(\frac{1}{2}\right)^{x_1} \left(\frac{\pi}{4}\right)^{x_2} \left(\frac{1-\pi}{4}\right)^{x_3} \left(\frac{1-\pi}{4}\right)^{x_4} \left(\frac{\pi}{4}\right)^{x_5}.$$

Observemos que en este caso la integral en $X(\mathbf{y})$ es sobre los pares $(x_1, x_2) = \{(0, 125), (1, 124), \dots, (125, 0)\}$, y que simplemente se sustituye (x_3, x_4, x_5) por $(18, 20, 34)$.

Al definir el algoritmo EM se mostrará cómo hallar $\pi^{(p+1)}$ a partir de $\pi^{(p)}$, donde $\pi^{(p)}$ denota el valor de π después de p iteraciones, para $p = 0, 1, 2, \dots$. El paso E, “*expectation*”, del algoritmo estima los estadísticos suficientes de los *datos completos* \mathbf{x} , dados los *datos observados* \mathbf{y} , en este caso. En general lo que hace el paso E es calcular la esperanza condicionada de la función $\log f(\mathbf{x} | \theta)$. En nuestro caso, (x_3, x_4, x_5) son conocidos, así que solamente tienen que ser

calculados los estadísticos suficientes para x_1 y x_2 donde $x_1 + x_2 = y_1 = 125$. Para la estimación de x_1 y x_2 usaremos el estimador correspondiente de π . Como la distribución es Multinomial, estas estimaciones, que serán las esperanzas condicionadas, tendrán las expresiones

$$\begin{aligned} x_1^{(p)} &= E[X_1 | \mathbf{Y}] = np_1 = 125 \frac{\frac{1}{2}}{\frac{1}{2} + \frac{\pi^{(p)}}{4}}, \\ x_2^{(p)} &= E[X_2 | \mathbf{Y}] = np_2 = 125 \frac{\frac{\pi^{(p)}}{4}}{\frac{1}{2} + \frac{\pi^{(p)}}{4}}. \end{aligned} \tag{1.2}$$

El paso M, “*maximization*”, considera los datos completos estimados $(x_1^{(p)}, x_2^{(p)}, 18, 20, 34)$ y calcula π por máxima verosimilitud como si los datos completos estimados fuesen los datos observados; entonces tendremos que

$$\pi^{(p+1)} = \frac{x_2^{(p)} + 34}{x_2^{(p)} + 34 + 18 + 20}, \tag{1.3}$$

puesto que

$$\log f(\mathbf{x} | \pi) \propto (x_2 + x_5) \log \frac{\pi}{4} + (x_3 + x_4) \log \frac{1 - \pi}{4},$$

y derivando respecto de π e igualando a cero se tiene que

$$\hat{\pi} = \frac{x_2 + x_5}{x_2 + x_5 + x_3 + x_4}$$

es el estimador máximo verosímil de π cuando se supone que todos los datos han sido observados.

Partiendo de un valor inicial $\pi^{(0)} = 0.5$, las primeras iteraciones del algoritmo están dadas en la tabla siguiente:

p	$\pi^{(p)}$	$\pi^{(p)} - \pi^*$	$(\pi^{(p+1)} - \pi^*)/(\pi^{(p)} - \pi^*)$
0	0.500000000	0.126821498	0.1465
1	0.608247423	0.018574075	0.1346
2	0.624321051	0.002500447	0.1330
3	0.626488879	0.000332619	0.1328
4	0.626777323	0.000044176	0.1328
5	0.626815632	0.000005866	0.1328
6	0.626820719	0.000000779	0.1328
7	0.626821395	0.000000104	0.1328
8	0.626821484	0.000000014	0.1328

Sustituyendo $x_2^{(p)}$ de la ecuación (1.2) en la ecuación (1.3), y tomando $\pi^* = \pi^{(p)} = \pi^{(p+1)}$ se obtiene el estimador de máxima-verosimilitud de π ($\pi^* = 0.6268214980$). La segunda columna de la tabla da los valores de las desviaciones $\pi^{(p)} - \pi^*$, la tercera columna da la razón de desviaciones sucesivas. Smith, en la discusión a Dempster et al. (1977), calcula el error estándar de π^* como el error estándar ordinario del estimador de una proporción. Se observa que, después de algunas iteraciones, el cociente $(\pi^{(p+1)} - \pi^*)/(\pi^{(p)} - \pi^*)$, que denotamos por λ , toma el valor constante 0.1328. Veremos posteriormente que este cociente es de gran utilidad para el cálculo de las Varianzas-Covarianzas entre los estimadores obtenidos (algoritmo SECM). El número total de individuos se puede estimar por

$$n^* = x_2^* + 34 + 18 + 20 = 125 \frac{\frac{\pi^*}{4}}{\frac{1}{2} + \frac{\pi^*}{4}} + 38 + 18 + 20 = 101.83.$$

Entonces, el error estándar de π^* es

$$\sqrt{\frac{\pi^*(1 - \pi^*)}{n^*(1 - \lambda)}} = 0.051,$$

que es el error estándar del estimador de una proporción con un factor de ponderación $(1 - \lambda)$ en el denominador.

1.1.2 Definición del Algoritmo EM

En este apartado se va a definir el algoritmo EM, según fue introducido por Dempster et al. (1977). Supongamos que la familia de densidades de los datos

completos es $f(\mathbf{x} | \theta)$ e introduzcamos una nueva función

$$Q(\theta' | \theta) = E[\log f(\mathbf{x} | \theta') | \mathbf{y}, \theta], \quad (1.4)$$

que se supondrá que existe para todo par (θ', θ) . En particular, se supondrá que $f(\mathbf{x} | \theta) > 0$ para casi todo elemento de \mathbf{X} y para todo $\theta \in \Theta$. Se define la iteración EM, $\theta^{(p)} \rightarrow \theta^{(p+1)}$, como sigue:

Paso E: Se calcula $Q(\theta | \theta^{(p)})$.

Paso M: Se elige un valor en Θ , $\theta^{(p+1)}$, que maximice $Q(\theta | \theta^{(p)})$.

La idea heurística es elegir θ^* que maximice $\log f(\mathbf{x} | \theta)$. Como no conocemos $\log f(\mathbf{x} | \theta)$ maximizamos su esperanza dados los *datos incompletos* \mathbf{y} y el correspondiente $\theta^{(p)}$. En el caso particular de familias exponenciales el paso E equivale a calcular sólo las esperanzas condicionadas de los estadísticos suficientes a \mathbf{y} y $\theta^{(p)}$.

1.1.3 Resultados de interés

Presentamos a continuación algunos resultados básicos aplicables al algoritmo EM dados por Dempster et al. (1977), a partir de ahora nos referiremos a este artículo como DLR, Boyles (1983) y Wu (1983). Supondremos que el observable \mathbf{y} es fijo y conocido.

La densidad de los *datos completos* puede ser factorizada como

$$f(\mathbf{x}, \mathbf{y} | \theta) = k(\mathbf{x} | \mathbf{y}, \theta)g(\mathbf{y} | \theta),$$

siendo $k(\mathbf{x} | \mathbf{y}, \theta)$ la densidad de los *datos completos* \mathbf{x} dados los *datos observados* \mathbf{y} . Hay que hacer notar que cuando la función $\mathbf{y}(\mathbf{x})$ es determinística, hecho que ocurre la mayoría de las veces (definir \mathbf{x} de forma adecuada), $f(\mathbf{x} | \theta) = f(\mathbf{x}, \mathbf{y} | \theta)$, con lo que el primer término lo tomaremos $f(\mathbf{x} | \theta)$. Si notamos $L(\theta) = \log g(\mathbf{y} | \theta)$ se puede observar que adopta una expresión del siguiente tipo en función de las densidades k y f

$$L(\theta) = \log f(\mathbf{x} | \theta) - \log k(\mathbf{x} | \mathbf{y}, \theta). \quad (1.5)$$

Además de la notación introducida anteriormente será conveniente escribir

$$H(\theta' | \theta) = E[\log k(\mathbf{x} | \mathbf{y}, \theta') | \mathbf{y}, \theta], \quad (1.6)$$

con lo que tomando esperanzas condicionadas en (1.5) resulta

$$Q(\theta' | \theta) = L(\theta') + H(\theta' | \theta).$$

Empleando esto y la desigualdad de Jensen (Rao (1972), pp. 47) se prueba el siguiente lema.

Lema 1.1.1 (DLR, p. 6)

Para cualquier par (θ', θ) en $\Theta \times \Theta$

$$H(\theta' | \theta) \leq H(\theta | \theta)$$

y se da la igualdad si y solamente si $k(\mathbf{x} | \mathbf{y}, \theta') = k(\mathbf{x} | \mathbf{y}, \theta)$ c.s.

La regla de iteración del algoritmo EM, aplicable para cualquier punto de partida, define una aplicación de Θ en Θ , $\theta \rightarrow M(\theta)$, tal que cada paso, un input $\theta^{(p)}$ nos genera un output $\theta^{(p+1)}$; es decir, tenemos una función M tal que

$$M(\theta^{(p)}) = \theta^{(p+1)}.$$

Definición 1.1.2

Un algoritmo iterativo con aplicación $M(\theta)$ es un algoritmo EM generalizado (abreviadamente GEM) si

$$Q(M(\theta) | \theta) \geq Q(\theta | \theta)$$

para todo $\theta \in \Theta$.

Observación 1.1.3

El algoritmo EM es un caso particular de un algoritmo GEM, ya que en la definición del algoritmo EM se necesitaba que

$$Q(M(\theta) | \theta) \geq Q(\theta' | \theta)$$

para cada par $(\theta', \theta) \in \Theta \times \Theta$, con lo que $\theta' = M(\theta)$ incrementa la función $Q(\theta' | \theta)$.

Teorema 1.1.4 (DLR, p. 7)*Para todo algoritmo GEM,*

$$L(M(\theta)) > L(\theta), \quad \forall \theta \in \Theta,$$

dándose la igualdad si y sólo si

$$Q(M(\theta) | \theta) = Q(\theta | \theta)$$

y

$$k(\mathbf{x} | \mathbf{y}, M(\theta)) = k(\mathbf{x} | \mathbf{y}, \theta), \quad c.s.$$

Consecuencias inmediatas del teorema anterior son los dos siguientes corolarios, éstos implican, que un estimador máximo verosímil de θ sea un punto fijo de un algoritmo GEM.

Corolario 1.1.5 (DLR, p. 7)

Supongamos que para $\theta^ \in \Theta$, $L(\theta^*) \geq L(\theta)$, para todo $\theta \in \Theta$. Entonces, para todo algoritmo GEM se verifica:*

1. $L(M(\theta^*)) = L(\theta^*)$.
2. $Q(M(\theta^*) | \theta^*) = Q(\theta^* | \theta^*)$.
3. $k(\mathbf{x} | \mathbf{y}, M(\theta^*)) = k(\mathbf{x} | \mathbf{y}, \theta^*)$ c.s.

Corolario 1.1.6 (DLR, p. 7)

Si existe $\theta^ \in \Theta$ tal que $L(\theta^*) > L(\theta)$, para todo $\theta \in \Theta$ distinto de θ^* , entonces para todo algoritmo GEM,*

$$M(\theta^*) = \theta^*.$$

Teorema 1.1.7 (DLR, p. 7, modificado por Boyles (1983))

Sea un algoritmo GEM, $\theta^{(p)}$, para $p=0,1,2,\dots$, tal que:

1. *La sucesión $\{L(\theta^{(p)})\}$ está acotada.*
2. *$Q(\theta^{(p+1)} | \theta^{(p)}) - Q(\theta^{(p)} | \theta^{(p)}) > \alpha \|\theta^{(p+1)} - \theta^{(p)}\|^2$, para algún $\alpha > 0$ y para todo p .*

Entonces, la sucesión $\{\theta^{(p)}\}$ cumple que

$$\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0, \quad \text{cuando } p \rightarrow \infty.$$

(Se supone que Θ es un espacio real r -dimensional y $\|\cdot\|$ es la norma euclídea).

El teorema 1.1.4 implica que $L(\theta)$ es creciente en cada iteración del algoritmo GEM, y si la función Q en la iteración $(p+1)$ aumenta (es decir, $Q(\theta^{(p+1)} | \theta^{(p)}) > Q(\theta^{(p)} | \theta^{(p)})$) es estrictamente creciente. Los corolarios implican que un estimador de máxima verosimilitud es un punto fijo de un algoritmo GEM. El teorema 1.1.7 da condiciones bajo las cuales

$$\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0, \quad \text{cuando } p \rightarrow \infty.$$

Estamos interesados en que el algoritmo GEM converja a un estimador de máxima verosimilitud bajo ciertas condiciones. Es natural introducir entre éstas, condiciones de continuidad y diferenciabilidad. Supondremos que Θ es una región dentro de un espacio real r -dimensional y la norma considerada es la euclídea, y supondremos la existencia y continuidad de un número suficiente de derivadas de las funciones $Q(\theta' | \theta)$, $L(\theta)$, $H(\theta' | \theta)$ y $M(\theta)$ para justificar los desarrollos en series de Taylor que serán usados posteriormente. También se supondrá que las operaciones de diferenciación y de tomar esperanzas se pueden intercambiar. Notaremos $Var[\dots]$ al operador de la covarianza condicionada.

Lema 1.1.8 (DLR, p. 8)

Para todo $\theta \in \Theta$,

$$E \left[\frac{\partial}{\partial \theta} \log k(\mathbf{x} | \mathbf{y}, \theta) | \mathbf{y}, \theta \right] = D^{10} H(\theta | \theta) = 0,$$

$$Var \left[\frac{\partial}{\partial \theta} \log k(\mathbf{x} | \mathbf{y}, \theta) | \mathbf{y}, \theta \right] = D^{11} H(\theta | \theta) = -D^{20} H(\theta | \theta).$$

Teorema 1.1.9 (DLR, p. 8, modificado por Boyles (1983))

Sea un algoritmo GEM, $\theta^{(p)}$, para $p=0,1,2,\dots$, tal que

$$D^{10} Q(\theta^{(p+1)} | \theta^{(p)}) = 0$$

Entonces para todo p , existe un $\theta_0^{(p+1)}$ sobre la línea del segmento que une $\theta^{(p)}$ y $\theta^{(p+1)}$ tal que

$$Q(\theta^{(p+1)} | \theta^{(p)}) - Q(\theta^{(p)} | \theta^{(p)}) = -(\theta^{(p+1)} - \theta^{(p)})' D^{20} Q(\theta_0^{(p+1)} | \theta^{(p)}) (\theta^{(p+1)} - \theta^{(p)}).$$

Además, si la sucesión $\{L(\theta^{(p)})\}$ está acotada, la sucesión $\{D^{20} Q(\theta_0^{(p+1)} | \theta^{(p)})\}$ es definida negativa y sus autovalores verifican la propiedad siguiente:

$(P_1) \exists \epsilon > 0 / \forall \lambda$ valor propio, se verifica que $\epsilon < -\lambda$.

En tal caso, la sucesión $\{\theta^{(p)}\}$ cumple que

$$\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0, \quad \text{cuando } p \rightarrow \infty.$$

Teorema 1.1.10 (DLR, p. 9)

Sea un algoritmo GEM, $\theta^{(p)}$, para $p = 0, 1, 2, \dots$, tal que

1. $\{\theta^{(p)}\}$ converge a θ^* en el cierre de Θ .
2. $D^{10} Q(\theta^{(p+1)} | \theta^{(p)}) = 0$.
3. $D^{20} Q(\theta^{(p+1)} | \theta^{(p)})$ es definida negativa y sus autovalores verifican la propiedad (P_1) .

Entonces,

1. $DL(\theta^*) = 0$.
2. $D^{20} Q(\theta^* | \theta^*)$ es definida negativa.
3. $DM(\theta^*) = [D^{20} Q(\theta^* | \theta^*)]^{-1} D^{20} H(\theta^* | \theta^*)$.

En el caso de una familia exponencial regular (notando por θ los parámetros naturales), se observa la siguiente relación

$$D^{20} Q(\theta | \theta^{(p)}) = -Var(\mathbf{t} | \theta).$$

Si los autovalores de $-Var(\mathbf{t} | \theta)$ verifican la propiedad (P_1) para algún $\theta_0^{(p+1)}$ en la línea que une $\theta^{(p)}$ y $\theta^{(p+1)}$, entonces

$$\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0, \quad \text{cuando } p \rightarrow \infty.$$



Observemos que en este caso,

$$D^{20}H(\theta^* | \theta^*) = -Var(t | y, \theta^*)$$

de donde

$$DM(\theta^*) = [Var(\mathbf{t} | \theta^*)]^{-1}Var(\mathbf{t} | y, \theta^*)$$

Según comentan Dempster et al. (1977), “En casi todas las aplicaciones, el límite θ^* de $\{\theta^{(p)}\}$ será un máximo local de $L(\theta)$, posiblemente global. Una excepción podría ocurrir si $DM(\theta^*)$ tuviese autovalores que excedieran a la unidad. En este caso θ^* podría ser un punto de silla de $L(\theta)$; para determinado $\{\theta^{(p)}\}$ convergente a θ^* , pudiera ocurrir que fuese ortogonal en el límite a los autovectores de $DM(\theta^*)$ asociados con los autovalores grandes. Observemos que, si a θ se le hiciera una pequeña perturbación aleatoria en torno a un punto de silla θ^* , entonces el algoritmo EM divergería a partir del punto de silla. Por tanto, generalmente esperamos que $DL^2(\theta^*)$ sea semidefinida negativa o definida negativa, en tales casos los autovalores de $DM(\theta^*)$ están en $[0,1]$ o en $[0,1)$, respectivamente. En vista de la igualdad, $D^{20}L(\theta^*) = D^{20}Q(\theta^* | \theta^*)(I - DM(\theta^*))$, un autovalor de $DM(\theta^*)$ que sea la unidad en un entorno de θ^* implica una cresta de $L(\theta)$ en θ^* ”.

Cuando los autovalores de $DM(\theta^*)$ son todos menores que uno, el autovalor más grande es el que da la velocidad de convergencia del algoritmo. La velocidad de convergencia depende directamente de los tamaños relativos de $D^2L(\theta^*)$ y $D^{20}H(\theta^* | \theta^*)$. Notemos que $D^2L(\theta^*)$ es una medida de la información de los datos y en torno a θ , mientras que $-D^{20}H(\theta^* | \theta^*)$ es la información esperada o de Fisher de la parte no observada de x en torno a θ . Así, si la información perdida debido a la incompletitud es pequeña el algoritmo convergerá rápidamente. La información perdida puede ser variable con respecto a las distintas componentes de θ ; por tanto, al aplicar el algoritmo EM puede ocurrir que, mientras algunas componentes de θ se aproximan rápidamente a sus correspondientes de θ^* , otras necesitan más iteraciones. Esto se explicará más detenidamente cuando presentemos los algoritmos suplementados.

Notación

Θ^0 denota el interior de Θ , y para números reales λ , $\{L \geq \lambda\}$ denota el conjunto $\{\theta \in \Theta : L(\theta) \geq \lambda\}$, de forma análoga se nota el conjunto $\{L = \lambda\}$, y $S = \{\theta \in \Theta^0 : DL(\theta) = 0\}$ denota al conjunto de puntos estacionarios de L . Vamos a suponer que $\lambda_0 = L(\theta^{(0)}) > -\infty$. Además, notaremos por

- (A) $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$, cuando $p \rightarrow \infty$.
- (B) $\{L \geq \lambda_0\}$ es compacto, para cualquier $L(\theta_0) > -\infty$.
- (C) L es continua en $\{L \geq \lambda_0\}$.
- (D) $\{L \geq \lambda_0\}$ es un subconjunto compacto de Θ^0 , para cualquier $L(\theta_0) > -\infty$.
- (E) L y $Q(\cdot | \theta)$, $\theta \in \{L \geq \lambda_0\}$, son diferenciables en un conjunto abierto que contiene $\{L \geq \lambda_0\}$.
- (F) M es continua en $\{L \geq \lambda_0\}$.
- (G) $D^{10}Q(M(\theta) | \theta) = 0$, para todo $\theta \in \{L \geq \lambda_0\}$.

La suposición (B) refleja nuestro interés en valores de los parámetros finitos. En algunos casos la no acotación del conjunto $\{L \geq \lambda_0\}$ puede evitarse con una elección apropiada de la parametrización.

Teorema 1.1.11 (Boyles (1983), p. 49)

Sea $\{\theta^{(p)}\}$ una sucesión GEM, y supongamos (A), (B) y (C). Entonces, existe $\lambda^* \in [\lambda_0, +\infty)$ tal que $L(\theta^{(p)}) \nearrow \lambda^*$ cuando $p \rightarrow \infty$. Además, $\{\theta^{(p)}\}$ converge a un compacto, componente conexa de $\{L = \lambda^*\}$.

Observación 1.1.12

Si $\{L = \lambda^*\}$ es un conjunto discreto, el teorema anterior implicará la convergencia de $\{\theta^{(p)}\}$ a algún $\theta^* \in \{L = \lambda^*\}$, cuando $p \rightarrow \infty$.

Teorema 1.1.13 (Boyles (1983), p. 49)

Sea $\{\theta^{(p)}\}$ una sucesión GEM, y supongamos (A), (D), (E), (F) y (G). Entonces, $\{\theta^{(p)}\}$ converge a un compacto, componente conexa de $S \cap \{L = \lambda^*\}$, donde λ^* está dado en el teorema anterior.

Observación 1.1.14

El teorema 1.1.13 no garantiza en sí mismo la convergencia de $\{\theta^{(p)}\}$ a un conjunto de máximos locales. En el caso particular que S sea un conjunto discreto implicará la convergencia de $\{\theta^{(p)}\}$ a algún $\theta^* \in S \cap \{L = \lambda^*\}$, cuando $p \rightarrow \infty$.

El propósito original del algoritmo EM era proporcionar un cálculo iterativo del estimador de máxima verosimilitud. Si para un algoritmo GEM la sucesión $\{L(\theta^{(p)})\}$ es acotada, se tiene la convergencia monótona a algún L^* .

Necesitamos saber si L^* es el máximo global de $L(\theta)$ en Θ , y en el caso en que no lo sea, si es un máximo local o un valor estacionario de L .

Como comentó Wu (1983), aunque en el paso M del algoritmo EM está involucrada una maximización global de Q , el término H en $L = Q - H$ puede no intervenir. Desafortunadamente, no siempre la convergencia es a un máximo local. Supongamos que $\{\theta^{(p)}\}$ convergiera a algún θ^* en el interior de Θ , que existieran las matrices hessianas $D^{20}Q(\theta^* | \theta^*)$ y $D^{20}H(\theta^* | \theta^*)$ con respecto a la primera variable, y que $D^{20}Q(\theta' | \theta)$ fuese continua en (θ', θ) . Entonces, por la definición del paso M, y por el lema 1.1.8, $-D^{20}Q(\theta^* | \theta^*)$ sería definida no negativa. Como $D^2L(\theta^*) = D^{20}Q(\theta^* | \theta^*) - D^{20}H(\theta^* | \theta^*)$, θ^* podría no ser un máximo local, con lo que no se aseguraría la convergencia a un máximo local.

Murray (1977) usando el algoritmo EM para estimar los parámetros de distribuciones normales multivariantes con muestras incompletas, observó que podría haber problemas en la aplicación del algoritmo cuando existía más de un valor estacionario. Dió un ejemplo en el que el algoritmo EM converge a un valor estacionario, pero no a un máximo local.

Definición 1.1.15

Una aplicación A de puntos de un conjunto X a subconjuntos de X se dice que es una aplicación punto a conjunto sobre X . Dicha función será cerrada en x , $x \in X$, si $x_k \rightarrow x$, donde $x_k \in X$, e $y_k \rightarrow y$, donde $y_k \in A(x_k)$, implica que $y \in A(x)$. En el caso particular que A sea una aplicación de punto a punto, la continuidad de A implica que A sea cerrada.

Teorema 1.1.16 (Teorema de Convergencia Global (Wu (1983), p. 97))

Sea una sucesión $\{x_k\}_{k=0,1,\dots}$ generada por $x_{k+1} \in M(x_k)$, donde M es una aplicación punto a conjunto sobre X , y sea un conjunto solución $\Gamma \subset X$. Supongamos que:

1. *Todos los puntos x_k están contenidos en un conjunto compacto $S \subset X$.*
2. *M es cerrada sobre el complemento de Γ .*
3. *Existe una función continua, α , sobre X tal que:*
 - (a) *Si $x \notin \Gamma$, $\alpha(y) > \alpha(x)$, $\forall y \in M(x)$.*

(b) Si $x \in \Gamma$, $\alpha(y) \geq \alpha(x)$, $\forall y \in M(x)$.

Entonces, todos los puntos límites de $\{x_k\}$ están en el conjunto solución Γ y $\{\alpha(x_k)\}$ converge monótonamente a $\alpha(x)$, para algún $x \in \Gamma$.

En nuestro caso, la aplicación M es la aplicación punto a punto de una iteración GEM y $\alpha(x)$ es la función de verosimilitud L . Tomaremos el conjunto solución Γ como uno de los dos siguientes conjuntos

\mathcal{M} =conjunto de máximos locales en el interior de Θ .

\mathcal{S} =conjunto de puntos estacionarios en el interior de Θ .

Supongamos (A), (B), (C) y (D), la condición $x \in \Gamma$, $\alpha(y) \geq \alpha(x)$, $\forall y \in M(x)$ se tiene claramente por la definición de la aplicación M en una iteración GEM, y la condición de que todos los puntos x_k están contenidos en un conjunto compacto $S \subset X$ se tiene de la suposición (B). Cuando sea necesario calcular las derivadas de L , Q y H en $\theta^{(p)}$ se supondrá que $\theta^{(p)}$ está en el interior de Θ . Esta suposición es consecuencia de que el conjunto $\{L \geq \lambda_0\}$ esté en el interior de Θ , donde $\lambda_0 = L(\theta^{(0)})$.

Teorema 1.1.17 (Wu (1983), p. 98)

Sea $\{\theta^{(p)}\}$ una sucesión GEM generada por $\theta^{(p+1)} \in M(\theta^{(p)})$, y supongamos que:

1. M es una aplicación cerrada punto a conjunto sobre el complemento de \mathcal{S} (respectivamente \mathcal{M}).
2. $L(\theta^{(p+1)}) > L(\theta^{(p)})$, $\forall \theta^{(p)} \notin \mathcal{S}$ (respectivamente \mathcal{M}).

Entonces, todos los puntos límites de $\{\theta^{(p)}\}$ son puntos estacionarios (respectivamente máximos locales) de L , y $\{L(\theta^{(p)})\}$ converge monótonamente a $L^* = L(\theta^*)$, para algún $\theta^* \in \mathcal{S}$ (respectivamente \mathcal{M}).

Para el caso de un algoritmo EM, una condición suficiente para que M sea cerrada es que

(H) $Q(\theta' | \theta)$ es continua en θ' y θ .

Esta condición es muy débil y será satisfecha en muchas situaciones prácticas.

Teorema 1.1.18 (Wu (1983), p. 98)

Supongamos que Q satisface la condición de continuidad (H). Entonces todos los puntos límites de cualquier sucesión $\{\theta^{(p)}\}$ de un algoritmo EM son puntos estacionarios de L y $\{L(\theta^{(p)})\}$ converge monótonamente a $L^* = L(\theta^*)$, para algún $\theta^* \in \mathcal{S}$.

El mismo argumento no es aplicable cuando \mathcal{S} se reemplaza por \mathcal{M} . Sea $\theta^{(p)} \in \mathcal{S} \setminus \mathcal{M}$; en este caso $DL(\theta^{(p)}) = D^{10}Q(\theta^{(p)} | \theta^{(p)}) = 0$, y $\theta^{(p)}$ puede realmente maximizar $Q(\theta | \theta^{(p)})$, donde $\theta \in \Theta$. La iteración EM termina en un punto estacionario pero no en un máximo local, y no se verifica que $L(\theta^{(p+1)}) > L(\theta^{(p)})$ para un tal $\theta^{(p)}$. Para garantizar la convergencia a un máximo local necesitamos imponer una condición adicional.

Teorema 1.1.19 (Wu (1983), p. 98)

Supongamos que Q satisface la condición de continuidad (H) y

$$\sup_{\theta' \in \Theta} Q(\theta' | \theta) > Q(\theta | \theta), \forall \theta \in \mathcal{S} \setminus \mathcal{M}.$$

Entonces todos los puntos límites de cualquier sucesión $\{\theta^{(p)}\}$ de un algoritmo EM son máximos locales de L y $\{L(\theta^{(p)})\}$ converge monótonamente a $L^ = L(\theta^*)$, para algún máximo local θ^* .*

Es difícil que la condición exigida en el teorema anterior se verifique, por lo que la utilidad del teorema será algo limitada. La convergencia de $\{L(\theta^{(p)})\}$ a L^* no implica automáticamente la convergencia de $\{\theta^{(p)}\}$ a un punto θ^* . (La implicación contraria es siempre cierta si $D^{10}H(\theta' | \theta)$ es continua). Esta última convergencia requiere condiciones más fuertes. Desde el punto de vista numérico, la convergencia de $\{\theta^{(p)}\}$ no es tan importante como la convergencia de $\{L(\theta^{(p)})\}$ a valores estacionarios o máximos locales, pero es interesante hacer un estudio de este problema.

Definimos $\mathcal{S}(a) = \{\theta \in \mathcal{S} : L(\theta) = a\}$, $\mathcal{M}(a) = \{\theta \in \mathcal{M} : L(\theta) = a\}$ y $L(a) = \{\theta \in \Theta : L(\theta) = a\}$. Bajo las condiciones del teorema 1.1.18, $L(\theta^{(p)}) \rightarrow L^*$ y todos los puntos límites de $\{\theta^{(p)}\}$ están en $\mathcal{S}(L^*)$ (respectivamente $\mathcal{M}(L^*)$). Si $\mathcal{S}(L^*)$ (respectivamente $\mathcal{M}(L^*)$) consta de un único punto θ^* , es decir, no puede haber dos puntos estacionarios distintos (respectivamente máximos locales) con el mismo valor L^* , entonces $\theta^{(p)} \rightarrow \theta^*$.

Teorema 1.1.20 (Wu (1983), p. 99)

Sea $\{\theta^{(p)}\}$ una sucesión de un algoritmo GEM que satisface las condiciones del teorema 1.1.17. Si $\mathcal{S}(L^) = \{\theta^*\}$ (respectivamente $\mathcal{M}(L^*) = \{\theta^*\}$), donde L^* es el límite de $\{L(\theta^{(p)})\}$ en el teorema 1.1.17, entonces $\theta^{(p)} \rightarrow \theta^*$.*

La suposición anterior se puede relajar si suponemos que $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$, cuando $p \rightarrow \infty$.

Teorema 1.1.21 (Wu (1983), p. 99)

Sea $\{\theta^{(p)}\}$ una sucesión de un algoritmo GEM que satisface las condiciones del teorema 1.1.17. Si $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$, cuando $p \rightarrow \infty$, entonces todos los puntos límites de $\{\theta^{(p)}\}$ están en un subconjunto conexo y compacto de $\mathcal{S}(L^*)$ (respectivamente $\mathcal{M}(L^*)$), donde L^* es el límite de $\{L(\theta^{(p)})\}$ en el teorema 1.1.17. En particular, si $\mathcal{S}(L^*)$ (respectivamente $\mathcal{M}(L^*)$) es discreto, es decir, sus únicas componentes conexas están aisladas, entonces $\{\theta^{(p)}\}$ converge a algún θ^* en $\mathcal{S}(L^*)$ (respectivamente $\mathcal{M}(L^*)$).

Observación 1.1.22

La segunda condición del teorema 1.1.17 para $\theta^{(p)} \notin \mathcal{S}$ se satisface automáticamente para cualquier sucesión EM. Obsérvese que los teoremas 1.1.11 y 1.1.13, conducen a las mismas conclusiones que el anterior pero bajo diferentes condiciones.

Teorema 1.1.23 (Wu (1983), p. 100)

Sea $\{\theta^{(p)}\}$ una sucesión de un algoritmo GEM que satisface la condición $D^{10}Q(\theta^{(p+1)} | \theta^{(p)}) = 0$. Supongamos que $D^{10}Q(\theta' | \theta)$ es continua en θ' y en θ . Si se verifica una de las siguientes condiciones:

1. $L(L^*) = \{\theta^*\}$.
2. $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$, cuando $p \rightarrow \infty$ y $L(L^*)$ es discreto.

Entonces $\{\theta^{(p)}\}$ converge a un punto estacionario θ^* con $L(\theta^*) = L^*$, el límite de $\{L(\theta^{(p)})\}$.

Observación 1.1.24

Notemos que para cualquier algoritmo EM se verifica que $D^{10}Q(\theta^{(p+1)} | \theta^{(p)}) = 0$. Por otra parte, puesto que $\mathcal{M}(a)$ y $\mathcal{S}(a)$ son subconjuntos de $L(a)$, las condiciones del teorema anterior son más fuertes que las correspondientes a los teoremas 1.1.20 y 1.1.21, respectivamente. La ventaja del teorema 1.1.23 es que no son necesarias las condiciones del teorema 1.1.17.

Corolario 1.1.25 (Wu (1983), p. 100)

Supongamos que $L(\theta)$ es unimodal en Θ con θ^* el único punto estacionario, y $D^{10}Q(\theta' | \theta)$ es continua en θ' y en θ . Entonces cualquier sucesión EM, $\{\theta^{(p)}\}$, converge al único punto θ^* que maximiza $L(\theta)$.

En la práctica el teorema 1.1.18 y el último corolario son los resultados más útiles ya que requieren condiciones que son muy fáciles de verificar. Las dos suposiciones claves son que el conjunto $\mathcal{S}(L^*)$ sea discreto y que $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$. En la práctica, la condición sobre la convergencia puede ser difícil de verificar. Por esto, se proponen condiciones suficientes, como la siguiente.

Condición

Existe una función forzante $\sigma(\cdot)$ tal que

$$Q(\theta^{(p+1)} | \theta^{(p)}) - Q(\theta^{(p)} | \theta^{(p)}) \geq \sigma(\|\theta^{(p+1)} - \theta^{(p)}\|), \quad \text{para todo } p.$$

(Una aplicación $\sigma : [0, \infty) \rightarrow [0, \infty)$ se dice ser una función forzante si para cualquier sucesión $\{t_k\}$ en $[0, \infty)$, la condición de que $\lim \sigma(t_k) = 0$, fuerza a que $\lim(t_k) = 0$).

Esta condición implica claramente que

$$\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0, \quad \text{cuando } p \rightarrow \infty.$$

Se sabe que

$$L(\theta^{(p+1)}) - L(\theta^{(p)}) \geq Q(\theta^{(p+1)} | \theta^{(p)}) - Q(\theta^{(p)} | \theta^{(p)}),$$

como $L(\theta^{(p)}) \rightarrow L^*$, y σ es una función forzante, se verifica que $\|\theta^{(p+1)} - \theta^{(p)}\| \rightarrow 0$.

1.2 Alisamiento y predicción de series temporales usando el algoritmo EM

1.2.1 Introducción

Datos de series temporales *incompletos* pueden originarse en problemas prácticos de varias formas. Un caso muy corriente es aquél en que las observaciones en el tiempo o en el espacio son obtenidas en intervalos irregulares; en general, es frecuente que las componentes de un vector puedan observarse sólo parcialmente. Otra situación de incompletitud es aquella en que se consideran datos que vienen afectados por un ruido y es imposible separar el ruido de la variable de interés. El problema general para estos modelos de series temporales es encontrar un estimador para la variable desconocida, sujeta a ciertas hipótesis

acerca de los parámetros asociados con el modelo que relaciona la variable con el ruido.

Una formulación muy general que parece englobar a una amplia clase de casos importantes es la modelización del espacio de estados, introducida por Kalman (1960) y Kalman y Bucy (1961). El modelo ha sido aplicado a la modelización de datos en muy diversas áreas científicas, particularmente experimentales y sociales.

Shumway y Stoffer (1982) proponen una forma de alisar y predecir una serie temporal haciendo uso del algoritmo EM y de la metodología del espacio de estados.

Supongamos que se observa un vector aleatorio, \mathbf{x}_t ($p \times 1$), indirectamente mediante el vector \mathbf{y}_t ($q \times 1$),

$$\mathbf{y}_t = \mathbf{M}_t \mathbf{x}_t + \mathbf{v}_t, \quad t = 1, \dots, n \quad (1.7)$$

donde \mathbf{v}_t ($q \times 1$) es un ruido gaussiano con matriz de covarianzas común \mathbf{R} ($q \times q$). Las matrices de diseño, \mathbf{M}_t ($q \times p$), son matrices conocidas que convierten los datos no observados, \mathbf{x}_t , en datos observados, \mathbf{y}_t . El vector aleatorio de interés es \mathbf{x}_t , que se supone satisface la ecuación

$$\mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \mathbf{w}_t, \quad t = 1, \dots, n \quad (1.8)$$

donde Φ_t es una matriz de transición $p \times p$, que describe cómo se comporta la serie en períodos sucesivos. La serie puede ser no estacionaria, ya que no se hacen suposiciones sobre las raíces de la ecuación característica. Se supone que el valor inicial \mathbf{x}_0 ($p \times 1$) es un vector aleatorio normal con media μ y matriz de covarianzas Σ ($p \times p$). Los vectores de ruido \mathbf{w}_t ($p \times 1$) para $t=1, \dots, n$, son normales, incorrelados, con media cero y matriz de covarianzas común \mathbf{O} , e independientes de \mathbf{v}_t . A la ecuación (1.7) se le suele llamar *ecuación de observación* y a la ecuación (1.8) *ecuación de estado*.

La formulación anterior no excluye la posibilidad de suponer un orden de autorregresión mayor para el vector \mathbf{x}_t , ya que cualquier modelo autorregresivo de orden r se puede reformular como un modelo autorregresivo de orden 1, aumentando la dimensión de la serie de la siguiente forma:

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-r+1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_r \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-r} \end{pmatrix} + \begin{pmatrix} \mathbf{w}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

y

$$\mathbf{y}_t = (\mathbf{M}_t, \mathbf{0}, \dots, \mathbf{0}) \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-r} \end{pmatrix} + \mathbf{v}_t$$

donde los $\mathbf{0}$ indican matrices o vectores formados por ceros, con las dimensiones correspondientes, e \mathbf{I} indica la matriz identidad $p \times p$.

Se trata de alisar y predecir la serie \mathbf{x}_t , o sea, se pretende estimar la serie no observada \mathbf{x}_t para $t = 1, \dots, n$ (alisamiento), y para $t = n + 1, n + 2, \dots$ (predicción), usando la serie observada $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$. Si uno conociese los valores de los parámetros μ , Σ , Φ , \mathbf{O} y \mathbf{R} , podría calcular los estimadores alisados de Kalman como esperanzas condicionadas; esto es equivalente a considerar \mathbf{x}_t como un vector paramétrico aleatorio que depende, en el sentido bayesiano, de los valores supuestos a priori para los parámetros.

Aplicando el algoritmo EM Shumway y Stoffer (1982) dan un método para el cálculo iterativo de los estimadores de máxima verosimilitud de los parámetros del modelo, y tratan el problema de alisado haciendo uso de la metodología del espacio de estados.

1.2.2 Estimación de los parámetros usando el algoritmo EM

Para la estimación de los parámetros del modelo de espacio de estados, definido por las ecuaciones (1.7)-(1.8), es necesario calcular el logaritmo de la función de verosimilitud conjunta de los datos completos $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n$, que puede ser escrita como

$$\begin{aligned} \log L \doteq & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{x}_0 - \mu)' \Sigma^{-1} (\mathbf{x}_0 - \mu) - \frac{n}{2} \log |\mathbf{O}| - \\ & - \frac{1}{2} \sum_{t=1}^n (\mathbf{x}_t - \Phi \mathbf{x}_{t-1})' \mathbf{O}^{-1} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1}) - \frac{n}{2} \log |\mathbf{R}| - \end{aligned}$$

$$-\frac{1}{2} \sum_{t=1}^n (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t)' \mathbf{R}^{-1} (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t) \tag{1.9}$$

donde $\log L$ debe ser maximizada respecto a los parámetros μ , Σ , Φ , \mathbf{O} y \mathbf{R} . La función de log-verosimilitud depende de la serie de datos no observados \mathbf{x}_t , $t = 0, 1, \dots, n$, por lo que no se pueden calcular los estimadores de los parámetros directamente y debe aplicarse otro procedimiento para el cálculo de los mismos. Shumway et al. (1982) trata el problema haciendo uso del algoritmo EM.

El paso E del algoritmo EM consistirá en calcular la esperanza condicionada de $\log L$ dados los datos observados $\mathbf{y}_1, \dots, \mathbf{y}_n$. Si los valores de los parámetros en la iteración r -ésima son $\mu^{(r)}$, $\Sigma^{(r)}$, $\Phi^{(r)}$, $\mathbf{O}^{(r)}$ y $\mathbf{R}^{(r)}$, se denotará por E_r a la esperanza condicionada relativa a una densidad con esos valores de los parámetros. Para calcular la esperanza condicionada de (1.9) se necesita calcular la media condicionada

$$\mathbf{x}_t^s = E_r[\mathbf{x}_t \mid \mathbf{y}_1, \dots, \mathbf{y}_s],$$

y las funciones de varianza-covarianza

$$\begin{aligned} \mathbf{P}_t^s &= \text{Var}_r[\mathbf{x}_t \mid \mathbf{y}_1, \dots, \mathbf{y}_s], \\ \mathbf{P}_{t,t-1}^s &= \text{Cov}_r[\mathbf{x}_t, \mathbf{x}_{t-1} \mid \mathbf{y}_1, \dots, \mathbf{y}_s]. \end{aligned}$$

\mathbf{x}_t^{t-1} representa el estimador del filtrado de Kalman, \mathbf{x}_t^n , $t = 0, 1, \dots, n$ es el estimador de alisamiento de \mathbf{x}_t con error cuadrático medio basado en todos los datos observados y \mathbf{x}_t^n , para $t > n$, es el valor predicho para la serie. Para el cálculo de \mathbf{x}_t^n y \mathbf{P}_t^n , se pueden emplear las siguientes recursiones basadas en resultados sobre el filtrado de Kalman, y además muestran un método recursivo para el cálculo de las covarianzas $\mathbf{P}_{t,t-1}^n$.

Observación 1.2.1

Otra forma de proceder sería particularizar el método que nosotros proponemos para procesos espaciales al caso de series temporales. Es decir, como se han impuesto condiciones de independencia y normalidad, en vez de calcular las esperanzas y varianzas-covarianzas condicionadas que necesitamos, mediante los calculos recursivos propuestos por el filtrado de Kalman, hacerlo mediante la ortogonalización y proyección de los datos incompletos. Este método se desarrollará en el capítulo tres.

El estimador de alisamiento de Kalman para el modelo definido por las ecuaciones (1.7) y (1.8), $(\mathbf{x}_t^n$ y $\mathbf{P}_t^n)$, se obtiene al minimizar el error cuadrático medio y se puede calcular recursivamente usando las ecuaciones siguientes (Jazwinski (1970)):

$$\begin{aligned}\mathbf{x}_t^{t-1} &= \Phi_t \mathbf{x}_{t-1}^{t-1} \\ \mathbf{P}_t^{t-1} &= \Phi_t \mathbf{P}_{t-1}^{t-1} \Phi_t' + \mathbf{O}_t \\ \mathbf{K}_t &= \mathbf{P}_t^{t-1} \mathbf{M}_t' (\mathbf{M}_t \mathbf{P}_t^{t-1} \mathbf{M}_t' + \mathbf{R}_t)^{-1} \\ \mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^{t-1}) \\ \mathbf{P}_t^t &= \mathbf{P}_t^{t-1} - \mathbf{K}_t \mathbf{M}_t \mathbf{P}_t^{t-1}\end{aligned}$$

para $t = 1, \dots, n$, donde $\mathbf{x}_0^0 = \mu$ y $\mathbf{P}_0^0 = \Sigma$. Para calcular \mathbf{x}_t^n y \mathbf{P}_t^n se realizan las recursiones siguientes,

$$\begin{aligned}\mathbf{J}_{t-1} &= \mathbf{P}_{t-1}^{t-1} \Phi_t' (\mathbf{P}_t^{t-1})^{-1} \\ \mathbf{x}_{t-1}^n &= \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1} (\mathbf{x}_t^n - \Phi_t \mathbf{x}_{t-1}^{t-1}) \\ \mathbf{P}_{t-1}^n &= \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1} (\mathbf{P}_t^n - \mathbf{P}_t^{t-1}) \mathbf{J}_{t-1}'\end{aligned}$$

para $t = n, n-1, \dots, 1$.

Para calcular $\mathbf{P}_{t,t-1}^n$ se usan las recursiones hacia atrás

$$\mathbf{P}_{t-1,t-2}^n = \mathbf{P}_{t-1}^{t-1} \mathbf{J}_{t-2}' + \mathbf{J}_{t-1} (\mathbf{P}_{t,t-1}^n - \Phi_t \mathbf{P}_{t-1}^{t-1}) \mathbf{J}_{t-2}'$$

para $t = n, n-1, \dots, 2$, donde

$$\mathbf{P}_{n,n-1}^n = (\mathbf{I} - \mathbf{K}_n \mathbf{M}_n) \Phi_n \mathbf{P}_{n-1}^{n-1}$$

Nota: Estas relaciones están deducidas en Shumway y Stoffer (1981).

Si se toman esperanzas condicionadas en (1.9) se tendrá

$$\begin{aligned}Q(\mu, \Sigma, \Phi, \mathbf{O}, \mathbf{R} | \mu^{(r)}, \Sigma^{(r)}, \Phi^{(r)}, \mathbf{O}^{(r)}, \mathbf{R}^{(r)}) &= E_r[\log L | \mathbf{y}_1, \dots, \mathbf{y}_n] = \\ &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \text{tr} \left(\Sigma^{-1} (\mathbf{P}_0^n + (\mathbf{x}_0 - \mu)(\mathbf{x}_0 - \mu)') \right) - \\ &= -\frac{n}{2} \log |\mathbf{O}| - \frac{1}{2} \text{tr} \left(\mathbf{O}^{-1} (\mathbf{C} - \mathbf{B} \Phi' - \Phi \mathbf{B}' + \Phi \mathbf{A} \Phi') \right) -\end{aligned}$$

$$-\frac{n}{2} \log |\mathbf{R}| - \frac{1}{2} \text{tr} \left(\mathbf{R}^{-1} \sum_{t=1}^n [(\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t)(\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t)' + \mathbf{M}_t \mathbf{P}_t^n \mathbf{M}_t'] \right),$$

donde tr denota la traza y

$$\mathbf{A} = \sum_{t=1}^n (\mathbf{P}_{t-1}^n + \mathbf{x}_{t-1}^n \mathbf{x}_{t-1}^{n'}),$$

$$\mathbf{B} = \sum_{t=1}^n (\mathbf{P}_{t,t-1}^n + \mathbf{x}_t^n \mathbf{x}_{t-1}^{n'}),$$

$$\mathbf{C} = \sum_{t=1}^n (\mathbf{P}_t^n + \mathbf{x}_t^n \mathbf{x}_t^{n'}).$$

Los términos \mathbf{x}_t^n , \mathbf{P}_t^n y $\mathbf{P}_{t,t-1}^n$ se calculan suponiendo los valores de los parámetros $\mu^{(r)}$, $\Sigma^{(r)}$, $\Phi^{(r)}$, $\mathbf{O}^{(r)}$ y $\mathbf{R}^{(r)}$, y usando las recursiones anteriores.

En el paso M del algoritmo se definirán los valores de los parámetros para la iteración (r+1)-ésima como aquellos valores de μ , Σ , Φ , \mathbf{O} y \mathbf{R} que maximizan la función $Q(\mu, \Sigma, \Phi, \mathbf{O}, \mathbf{R} | \mu^{(r)}, \Sigma^{(r)}, \Phi^{(r)}, \mathbf{O}^{(r)}, \mathbf{R}^{(r)})$. Empleando las reglas de derivación matricial se obtienen los siguientes valores de los parámetros,

$$\Phi^{(r+1)} = \mathbf{B} \mathbf{A}^{-1} \quad (1.10)$$

$$\mathbf{O}^{(r+1)} = n^{-1} (\mathbf{C} - \mathbf{B} \mathbf{A}^{-1} \mathbf{B}') \quad (1.11)$$

$$\mathbf{R}^{(r+1)} = n^{-1} \sum_{t=1}^n [(\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^n)(\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^n)' + \mathbf{M}_t \mathbf{P}_t^n \mathbf{M}_t'] \quad (1.12)$$

El primer término de la función $Q(\mu, \Sigma, \Phi, \mathbf{O}, \mathbf{R} | \mu^{(r)}, \Sigma^{(r)}, \Phi^{(r)}, \mathbf{O}^{(r)}, \mathbf{R}^{(r)})$ corresponde al de una normal multivariante con una única réplica, se puede tomar $\mu^{(r+1)} = \mathbf{x}_0^n$ y se fija el valor de Σ , de alguna forma razonable. Si \mathbf{x}_0 se considera como un valor fijo se elige $\Sigma=0$.

El valor de la función de log-verosimilitud en cada iteración puede calcularse usando la forma de las innovaciones (Gupta y Mehra (1974))

$$\begin{aligned} \log L \doteq & -\frac{1}{2} \sum_{t=1}^n \log |\mathbf{M}_t \mathbf{P}_t^{t-1} \mathbf{M}_t' + \mathbf{R}_t| - \\ & -\frac{1}{2} \sum_{t=1}^n (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^{t-1})' (\mathbf{M}_t \mathbf{P}_t^{t-1} \mathbf{M}_t' + \mathbf{R}_t)^{-1} (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^{t-1}) \end{aligned} \quad (1.13)$$

El procedimiento recomendado para llevar a cabo los cálculos sería el siguiente:

1. Suponer los valores iniciales de los parámetros $\mu^{(0)}$, $\Phi^{(0)}$, $\mathbf{O}^{(0)}$ y $\mathbf{R}^{(0)}$ y fijar Σ .
2. Calcular \mathbf{x}_t^n , \mathbf{P}_t^n y $\mathbf{P}_{t,t-1}^n$ usando el proceso iterativo anterior.
3. Evaluar la función de log-verosimilitud (1.13).
4. Estimar $\mu^{(1)} = \mathbf{x}_0^n$, y usar las ecuaciones (1.10) – (1.11) – (1.12) para hallar los valores $\Phi^{(1)}$, $\mathbf{O}^{(1)}$ y $\mathbf{R}^{(1)}$.
5. Volver al paso 1 hasta que los estimadores y la función de log-verosimilitud (1.13) sean estables.

1.3 El algoritmo ECM

El paso M del algoritmo EM consiste en el cálculo del máximo de una función matricial, $Q(\theta'|\theta)$. Si la función anterior no tiene una expresión simple, este cálculo puede ser bastante engorroso e incluso puede hacer que el número de iteraciones de los procedimientos numéricos que se utilicen sea comparativamente muy elevado con el número de iteraciones necesarias para encontrar el estimador de máxima verosimilitud mediante el algoritmo EM. Por esto, en los últimos años se han desarrollado algunas modificaciones del algoritmo EM que simplifican en gran medida el paso M del algoritmo. Una de estas modificaciones es el algoritmo ECM. Este algoritmo ha sido recientemente desarrollado y fue introducido por Meng y Rubin (1993). Básicamente este algoritmo es prácticamente igual que el algoritmo EM salvo que el paso de maximización lo divide en una secuencia de S pasos de maximización condicionados. En cada paso se maximiza la función Q sobre θ pero fijando alguna función vectorial de θ en su valor previo, que notaremos $g_s(\theta)$, ($s = 1, \dots, S$). Hay que hacer señalar que el resultado de estos S pasos, no tiene por que coincidir, y de hecho frecuentemente no coincide, con el valor de los parámetros que hacen máxima a la función Q en cada iteración del algoritmo EM. Sin embargo, ya que en cada paso se incrementa la función Q , como en el algoritmo EM se incrementa monótonamente la verosimilitud de θ . Más aún, cuando el conjunto de funciones g_s nos proporcione un conjunto rellenante del espacio (*space-filling*), en

el sentido que nos da la maximización sin restricciones sobre θ en su espacio paramétrico, el algoritmo ECM converge a un punto estacionario bajo esencialmente las mismas condiciones que garantizan la convergencia del algoritmo EM.

Introduzcamos formalmente el algoritmo ECM. Como hemos dicho con anterioridad, las maximizaciones se hacen condicionadas a unas funciones g_s en cada paso. Al conjunto formado por todas estas funciones vectoriales de θ seleccionadas con anterioridad lo vamos a notar por G . Es decir:

$$G = \{g_s(\theta); s = 1, \dots, S\}.$$

Comenzando con una condición inicial $\theta^{(0)} \in \Theta$ en la $(p + 1)$ -ésima iteración, $p = 0, 1, \dots$, el algoritmo ECM en primer lugar realiza el paso E igual que lo hacía el algoritmo EM y posteriormente desdobra el paso de maximización en S pasos de maximización condicionados (pasos CM) definidos como siguen: para $s = 1, \dots, S$, encontrar $\theta^{(p+s/S)}$ de tal forma que maximice la función $Q(\theta/\theta^{(p)})$ sobre $\theta \in \Theta$ sujeto a la restricción de que $g_s(\theta) = g_s(\theta^{(p+(s-1)/S)})$. Es decir, para $s = 1, \dots, S$, el s -ésimo paso CM en la p -ésima iteración del ECM consiste en encontrar $\theta^{(p+s/S)}$ tal que

$$\begin{aligned} Q(\theta^{(p+s/S)}/\theta^{(p)}) &\geq Q(\theta/\theta^{(p)}) \\ \forall \theta \in \Theta_s(\theta^{(p+(s-1)/S)}) &\equiv \{\theta \in \Theta : g_s(\theta) = g_s(\theta^{(p+(s-1)/S)})\}. \end{aligned} \tag{1.14}$$

Por tanto, el valor de θ para comenzar la siguiente iteración del ECM, $\theta^{(p+1)}$, se define como la salida del paso final S -ésimo, es decir, $\theta^{(p+S/S)}$.

Definición 1.3.1

Un algoritmo iterativo es llamado un algoritmo ECM si la $(p + 1)$ -ésima iteración comienza con un paso E (cálculo de esperanzas condicionadas a la información observada) para calcular $Q(\theta/\theta^{(p)})$ como función de θ , y es seguida de S (≥ 1) pasos CM, en cada uno de los cuales se encuentra $\theta^{(p+s/S)}$ como en (1.14), para $s = 1, \dots, S$.

Para garantizar que la convergencia del ECM sea tan apropiada como la del EM, hay que exigir ciertas restricciones sobre el conjunto de funciones, G , para que el máximo resultante en la última iteración, sea el máximo de

L en Θ . Esto puede lograrse requiriendo que G proporcione un conjunto rellanante del espacio como se precisará en la siguiente definición.

Definición 1.3.2

Sea $T_s(\theta)$ ($s = 1, \dots, S$) el conjunto de todas las direcciones posibles en $\theta \in \Theta$ con respecto al espacio de restricciones

$$\Theta_s(\theta) = \{\zeta \in \Theta : g_s(\zeta) = g_s(\theta)\},$$

esto es,

$$T_s(\theta) = \left\{ \eta \in \mathbb{R}^r : \exists \{\theta_n\} \subset \Theta_s(\theta) \text{ tal que } \eta = \lim_{n \rightarrow \infty} \frac{\theta_n - \theta}{\|\theta_n - \theta\|} \right\}.$$

Decimos que $G = \{g_s(\theta); s = 1, \dots, S\}$ es un conjunto rellanante del espacio en $\theta \in \Theta$ si

$$T(\theta) \equiv \text{clausura} \left\{ \sum_{s=1}^S a_s \eta_s : a_s \geq 0, \eta_s \in T_s(\theta) \right\} = \mathbb{R}^d. \quad (1.15)$$

En la literatura de optimización, $T_s(\theta)$ es llamado como ‘cono tangente’ ya que $\eta \in T_s(\theta)$ implica que $\alpha\eta \in T_s(\theta)$ para cualquier $\alpha \geq 0$. La intuición nos dice que (1.15) nos asegura que en cualquier punto de Θ podemos buscar el máximo en cualquier dirección, así que el resultado de la maximización no está restringido a un subespacio de Θ .

Supongamos que las funciones $g_s(\theta)$ ($s = 1, \dots, S$) son diferenciables y que su correspondiente gradiente, $\nabla g_s(\theta)$, es de rango máximo en $\theta \in \Theta_0$, el interior de Θ . Se puede ver que la condición (1.15) es equivalente a:

$$J(\theta) \equiv \bigcap_{s=1}^S J_s(\theta) = \{0\}, \quad (1.16)$$

donde $J_s(\theta)$ es el espacio generado por el gradiente, es decir,

$$J_s(\theta) = \{\nabla g_s(\theta)\lambda : \lambda \in \mathbb{R}^{r_s}\}$$

y r_s es la dimensión de la función vectorial $g_s(\theta)$. La intuición nos presenta esta equivalencia como válida ya que al buscar en espacios de la forma $g_s(\theta) = k$, es decir, en curvas de nivel, el gradiente nos define un subespacio normal a las

direcciones en las que se busca, por tanto, las direcciones donde no se busca en la iteración s . Por ello, buscar en todas las direcciones que es la condición (1.15) equivale a que el conjunto de las direcciones donde no se busca sea el $\{0\}$ que nos da la condición (1.16). Más formalmente se puede ver que (1.16) es una consecuencia directa de la siguiente identidad,

$$J(\theta) = \{\zeta : \zeta^T \eta \leq 0 \text{ para todo } \eta \in T(\theta)\},$$

lo cual se deduce directamente de los teoremas polar y bipolar en la literatura de optimización con restricciones. La ventaja de la expresión (1.16) sobre (1.15) es que se puede verificar directamente en muchas aplicaciones. El algoritmo EM se puede ver como un caso particular de algoritmo ECM con $S = 1$ y $g_1(\theta) \equiv \text{constante}$, la condición (1.16) es satisfecha automáticamente ya que $\nabla g_1(\theta) = 0$ para todo θ .

Sobre la convergencia del algoritmo ECM vamos a presentar dos resultados que aparecen en la página 275 del artículo de Meng et al. (1993), que bajo las mismas condiciones exigidas sobre el algoritmo EM en Wu (1983), nos aseguran la convergencia de este algoritmo.

Teorema 1.3.3 (Meng et al. (1993), p. 275)

Supongamos que todas las maximizaciones condicionadas de (1.14) son únicas. Entonces todos los puntos límites de cualquier sucesión ECM $\theta^{(p)}$ son puntos estacionarios de $L(\theta|Y)$ si G es un espacio relleno en todo $\theta^{(p)}$.

Teorema 1.3.4 (Meng et al. (1993), p. 275)

Supongamos que $L(\theta|Y)$ es unimodal en $\theta \in \Theta$ con único punto estacionario θ^ . Entonces cualquier sucesión ECM $\{\theta^{(p)}\}$ converge al único máximo θ^* si G es un espacio relleno en todo $\theta^{(p)}$, y ocurre o (a) cada maximización CM es única o (b) $D^{10}Q(\theta|\theta')$ es continua en ambos parámetros, θ y θ' , y $\nabla g_s(\theta)$ es continua en θ para $s = 1, \dots, S$.*

Otra de las variantes del algoritmo EM es el algoritmo MCECM (MC de multi-ciclo). Como se puede ver en la construcción del algoritmo ECM lo que se hace es calcular en un primer paso la esperanza condicionada y después S pasos de maximización condicionada. Podría pensarse en no mantener la misma esperanza condicionada en todos los pasos de maximización condicionada, sino ir renovándola cada cierto número de iteraciones, o en cada iteración. Este es el algoritmo MCECM.

1.4 Algoritmos suplementados

Los algoritmos suplementados (EM, ECM, MCECM) no son más que algoritmos en los que además de dar las estimaciones de los parámetros, se “suplementan” éstas, con una matriz de varianzas-covarianzas asintótica para ellos. A estos algoritmos les añadiremos una S delante para indicar que son suplementados. En el resto del capítulo vamos a utilizar la siguiente notación. Suponemos que los datos completos se pueden descomponer de la siguiente forma $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, donde \mathbf{Y} representa la parte observada de \mathbf{X} y \mathbf{Z} la parte no observada (*missing*). Así en nuestro esquema los datos incompletos $\mathbf{Y} = \mathbf{y}(\mathbf{X})$, consisten en sólo las primeras variables de los datos completos.

El algoritmo que utilicemos define implícitamente una función $\theta \rightarrow M(\theta)$ del espacio paramétrico, Θ , en si mismo tal que

$$\theta^{(p+1)} = M(\theta^{(p)}), \quad \text{para } p = 0, 1, \dots$$

Si $\theta^{(p)}$ converge a algún θ^* y $M(\theta)$ es continua, entonces θ^* debe de satisfacer que

$$\theta^* = M(\theta^*).$$

Más aún, si M es diferenciable en un entorno de θ^* , por la expansión de Taylor tenemos que

$$\theta^{(p+1)} - \theta^* \approx (\theta^{(p)} - \theta^*)DM, \quad (1.17)$$

donde

$$DM = \left(\frac{\partial M_j(\theta)}{\partial \theta_i} \right) \Big|_{\theta=\theta^*}$$

es la matriz jacobiana $r \times r$ para $M(\theta) = (M_1(\theta), \dots, M_r(\theta))$ evaluada en $\theta = \theta^*$. Así el algoritmo es esencialmente, en un entorno de θ^* , una iteración lineal con matriz de “tasa” DM , ya que generalmente DM será no idénticamente nula. Por esta razón, $DM(\theta^*)$ a menudo se llama matriz de la tasa de convergencia, o simplemente tasa de convergencia.

Es bien conocido (Little et al. (1987), pp.145) que para muestras grandes la matriz de varianzas-covarianzas de $(\theta - \theta^*)$ basada en \mathbf{Y} , V , puede encontrarse como la inversa de la matriz de información observada,

$$V = \mathbf{I}_o^{-1}(\theta^*|\mathbf{Y}), \quad (1.18)$$

donde $\mathbf{I}_o(\theta|\mathbf{Y})$ viene dada por la expresión:

$$\mathbf{I}_o(\theta|\mathbf{Y}) = -\frac{\partial^2 \log f(\mathbf{Y}|\theta)}{\partial \theta^2}. \quad (1.19)$$

Esta función es difícil de evaluar directamente. Sin embargo, la matriz de información de los datos completos observados,

$$\mathbf{I}_o(\theta|\mathbf{X}) = -\frac{\partial^2 \log f(\mathbf{X}|\theta)}{\partial \theta^2}, \quad (1.20)$$

cuya inversa nos da la matriz de varianzas-covarianzas de los datos completos, es frecuentemente una función simple, así como su esperanza respecto de la distribución condicional $f(\mathbf{Z}|\mathbf{Y}, \theta)$ evaluada en $\theta = \theta^*$:

$$\mathbf{I}_{oc} = \mathbf{E}[\mathbf{I}_o(\theta|\mathbf{X}) | \mathbf{Y}, \theta]_{\theta=\theta^*}. \quad (1.21)$$

La relación que existe entre estas matrices se pone de manifiesto en el siguiente desarrollo

$$f(\mathbf{Y}|\theta)f(\mathbf{Z}|\mathbf{Y}, \theta) = f(\mathbf{Y}, \mathbf{Z}|\theta) = f(\mathbf{X}|\theta),$$

tomando logaritmos y despejando la primera densidad obtenemos

$$\log(\mathbf{Y}|\theta) = \log f(\mathbf{X}|\theta) - \log f(\mathbf{Z}|\mathbf{Y}, \theta),$$

si realizamos $-\frac{\partial^2}{\partial \theta^2}$ obtenemos la igualdad

$$\mathbf{I}_o(\theta|\mathbf{Y}) = \mathbf{I}_o(\theta|\mathbf{X}) + \frac{\partial^2 \log f(\mathbf{Z}|\mathbf{Y}, \theta)}{\partial \theta^2},$$

tomando esperanzas condicionadas respecto de la distribución condicional $f(\mathbf{Z}|\mathbf{Y}, \theta)$ evaluada en $\theta = \theta^*$

$$\mathbf{I}_o(\theta^*|\mathbf{Y}) = \mathbf{I}_{oc} - \mathbf{I}_{om}, \quad (1.22)$$

donde

$$\mathbf{I}_{om} = \mathbf{E} \left[-\frac{\partial^2 \log f(\mathbf{Z}|\mathbf{Y}, \theta)}{\partial \theta^2} | \mathbf{Y}, \theta \right]_{\theta=\theta^*}.$$

Hay que observar que el primer miembro de (1.22) no cambia al tomar esperanzas, ya que la esperanza condicionada verifica que $\mathbf{E}[h(\mathbf{U}) | \mathbf{U}] = h(\mathbf{U})$,

siendo \mathbf{U} multidimensional. Así,

$$\mathbf{I}_o(\theta^*|\mathbf{Y}) = \mathbf{E}[\mathbf{I}_o(\theta|\mathbf{Y}) | \mathbf{Y}, \theta]_{\theta=\theta^*}$$

En DLR p. 9, la relación que aparece como (3.19) nos muestra que

$$DM(\theta^*) = DM^{EM}(\theta^*) = \mathbf{I}_{om} \mathbf{I}_{oc}^{-1}.$$

Por ello

$$\mathbf{I}_o(\theta^*|\mathbf{Y}) = \mathbf{I}_{oc} - \mathbf{I}_{om} = (\mathbf{I} - \mathbf{I}_{om} \mathbf{I}_{oc}^{-1}) \mathbf{I}_{oc} = (\mathbf{I} - DM^{EM}) \mathbf{I}_{oc}$$

Por tanto, para el algoritmo EM la matriz de varianzas-covarianzas asintótica adopta una expresión de la forma

$$V = \mathbf{I}_{oc}^{-1} (\mathbf{I} - DM^{EM})^{-1} \quad (1.23)$$

donde \mathbf{I} es la matriz identidad $r \times r$. Así, para el cálculo de la matriz V hay que calcular la matriz \mathbf{I}_{oc} , la matriz DM y realizar los cálculos de la expresión (1.23).

El cálculo de \mathbf{I}_{oc} es simplemente calcular una matriz que contenga las derivadas $\frac{\partial^2}{\partial \theta_i \partial \theta_j}$ y después calcular $\frac{r(r+1)}{2}$ esperanzas condicionadas (la matriz es simétrica) equivalentes al paso E del algoritmo y evaluarlas en $\theta = \theta^*$. La complejidad de este cálculo depende, obviamente, de la expresión que adopta la densidad de los datos completos \mathbf{X} . En el caso que sea una distribución de la familia exponencial se puede ver (Meng et al. (1991)) que si $S(\mathbf{X})$ es un estadístico suficiente y $S^*(\mathbf{Y}) = \mathbf{E}[S(\mathbf{X})|\mathbf{Y}, \theta^*]$, entonces

$$\mathbf{I}_{oc} = \mathbf{I}_o(\theta^*|S^*(\mathbf{Y})).$$

Para el cálculo de DM no podemos usar directamente la aproximación (1.17) ya que sólo obtendríamos algunos de los autovalores (en algunos casos sólo el mayor) y no la matriz completa. De la definición de derivada parcial podemos construir un método para dicho cálculo

$$\begin{aligned} r_{ij} &= \frac{\partial M_j(\theta^*)}{\partial \theta_i} = \\ &= \lim_{\theta_i \rightarrow \theta_i^*} \frac{M_j(\theta_1^*, \dots, \theta_{i-1}^*, \theta_i, \theta_{i+1}^*, \dots, \theta_d^*) - M_j(\theta^*)}{\theta_i - \theta_i^*} = \\ &= \lim_{p \rightarrow \infty} \frac{M_j(\theta^{(p)}(i)) - \theta_j^*}{\theta_i^{(p)} - \theta_i^*} \equiv \lim_{p \rightarrow \infty} r_{ij}^{(p)}, \end{aligned} \quad (1.24)$$

donde las definiciones intermedias se identifican claramente.

Por esto, una iteración del algoritmo SEM podría ser la siguiente:

ENTRADAS: θ^* y $\theta^{(p)}$.

1. Ejecutar los pasos usuales E y M para obtener $\theta^{(p+1)}$.
2. Calcular $\theta^{(p)}(i)$ de (1.24) y ejecutar una iteración del algoritmo EM para calcular $\tilde{\theta}^{(p+1)}(i)$.
3. Obtener los cocientes

$$r_{ij}^{(p)} = \frac{\tilde{\theta}_j^{(p+1)}(i) - \theta_j^*}{\theta_i^{(t)} - \theta_i^*}, \quad \text{para } j = 1, \dots, r. \quad (1.25)$$

4. Repetir los pasos 2 y 3 para $i = 1, \dots, r$.

SALIDAS: $\theta^{(p+1)}$ y $\{r_{ij}^{(p)} : i, j = 1, \dots, r\}$.

Como se puede observar, uno de los datos que hay que introducir es θ^* , el cual se puede calcular, por ejemplo, ejecutando el algoritmo EM hasta que se logre convergencia. Si esto se hace, podemos almacenar, además de θ^* , los resultados que propone como estimadores en cada una de las iteraciones y nos ahorraríamos el paso 1 del algoritmo anterior. Así, sólo nos quedaría hacer una iteración del algoritmo EM cambiando la condición inicial por $\theta_i^{(t)}$. Por tanto, si hemos tenido que realizar T iteraciones para estimar un parámetro de dimensión r , deberemos realizar $T \times r$ iteraciones cambiando en cada una de ellas la condición inicial y calcular los cocientes dados por (1.25) en cada iteración.

Se puede también aproximar DM sin obtener primero θ^* . La idea básica es aproximar r_{ij} por

$$\tilde{r}_{ij}^{(p)} = \frac{M_j(\theta_1^{(p-1)}, \dots, \theta_{i-1}^{(p-1)}, \theta_i^{(p)}, \theta_{i+1}^{(p-1)}, \dots, \theta_d^{(p-1)}) - \theta_j^{(p)}}{\theta_i^{(p)} - \theta_i^{(p-1)}} \quad (1.26)$$

más bien que por $r_{ij}^{(p)}$ de (1.25). Sin embargo este cálculo es bastante más inestable, sobre todo si en una sucesión uno de los parámetros está muy cerca del estimador final mientras que otro aún está relativamente lejos del suyo.

Luego ya tenemos un método para construir el algoritmo SEM, o bien directamente si disponemos de θ^* o bien ejecutando una vez el algoritmo EM

y después calculando sus covarianzas como se ha indicado anteriormente. Para los algoritmos SECM y SMCECM podemos buscar una relación directa entre la matriz de covarianzas y matrices que se obtengan directamente de estos algoritmos (sin S), o encontrar una relación entre la tasa de convergencia del algoritmo EM (DM^{EM}) y las de los algoritmos ECM (DM^{ECM}) y MCECM (DM^{MCECM}). Esto último aparece en el artículo de Meng (1994). Hay que hacer notar que en todos los casos la matriz \mathbf{I}_{oc} no depende del algoritmo que se emplee, sino tan sólo de la condición final θ^* .

Teorema 1.4.1 (Meng (1994))

Supongamos que todas las salidas de un algoritmo ECM, $\theta^{(p+s/S)}$ ($p \geq 1, \dots, S$) satisfacen las ecuaciones del multiplicador de Lagrange para maximización con restricciones, y $\theta^{(p+s/S)} \rightarrow \theta^*$ cuando $p \rightarrow \infty$ para $s = 1, \dots, S$. Entonces la tasa de convergencia del ECM viene dada por

$$DM^{ECM}(\theta^*) = DM^{EM}(\theta^*) + [\mathbf{I} - DM^{EM}(\theta^*)] \prod_{s=1}^S P_s, \quad (1.27)$$

donde

$$P_s = \nabla_s [\nabla'_s \mathbf{I}_{oc}^{-1}(\theta^*) \nabla_s]^{-1} \nabla'_s \mathbf{I}_{oc}^{-1}(\theta^*), \quad s = 1, \dots, S \quad (1.28)$$

con $\nabla_s = \nabla g_s(\theta^*)$ y $\prod_{s=1}^S P_s \equiv P_1 \cdots P_s$.

Supongamos que en el algoritmo MCECM insertamos un paso E antes del s_k -ésimo paso CM ($k = 1, \dots, K \leq S, s_1 = 1$). Entonces este algoritmo consiste en:

$$Q(\theta^{(p+s/S)} | \theta^{(p+(s_k-1)/S)}) \geq Q(\theta | \theta^{(p+(s_k-1)/S)}),$$

para todo $\theta \in \Theta_s(\theta^{(p+(s-1)/S)})$; $s_k \leq s < s_{k+1}$. (1.29)

Teorema 1.4.2 (Meng (1994))

Bajo las condiciones del teorema 1.4.1, la tasa de convergencia de un algoritmo MCECM (con K ciclos) dado por (1.29) es

$$DM^{MCECM}(\theta^*) = \prod_{k=1}^K \{DM^{EM}(\theta^*) + [\mathbf{I} - DM^{EM}(\theta^*)] \prod_{s=s_k}^{s_{k+1}-1} P_s\}, \quad (1.30)$$

donde $s_1 \equiv 1$ y $s_{K+1} \equiv S + 1$.

Notemos que en ausencia de observaciones omitidas, es decir cuando $\mathbf{X} = \mathbf{Y}$, la matriz $DM^{EM} = 0$ con lo que la ecuación (1.27) implica que

$$DM^{CM}(\theta^*) = \prod_{s=1}^S P_s, \quad (1.31)$$

donde M^{CM} es la función descrita por los pasos de maximización condicionada. Por tanto, de (1.27) y (1.31) podemos deducir que

$$(\mathbf{I} - DM^{ECM}(\theta^*)) = (\mathbf{I} - DM^{EM}(\theta^*))(\mathbf{I} - DM^{CM}(\theta^*)). \quad (1.32)$$

Observemos que en la ecuación (1.23) aparece la matriz DM^{EM} por lo que en el caso del algoritmo ECM su matriz de varianzas-covarianzas asintótica tendrá la expresión

$$V = \mathbf{I}_{oc}^{-1}(\mathbf{I} - DM^{CM}(\theta^*))(\mathbf{I} - DM^{ECM}(\theta^*))^{-1} \quad (1.33)$$

Las matrices $\mathbf{I} - DM^*$ son interpretadas como matriz de velocidad de convergencia de \star , ya que por ejemplo si tomamos el mayor autovalor de DM^* y éste está próximo a 1, indicará que la convergencia es lenta (la función M varía bastante) por lo que el menor autovalor, s , de la matriz que proponemos como velocidad de convergencia, nos indicará que si s es bajo su velocidad será baja y si es alto su velocidad será alta. Así la ecuación (1.32) nos indica que la matriz velocidad de convergencia del algoritmo ECM se puede calcular como el producto de las matrices de convergencia de los algoritmos EM y CM, lo cual parece lógico, ya que el algoritmo ECM se puede ver como la composición de los algoritmos EM y CM.

Revisamos la ecuación (1.33). Para calcular la matriz de covarianzas del algoritmo ECM necesitamos calcular \mathbf{I}_{oc} que se calcula independientemente del algoritmo empleado, DM^{ECM} que se obtendrá mediante la ecuación (1.25), teniendo claramente en cuenta que en el paso 2 de este procedimiento una iteración son S pasos CM, y por último calcular DM^{CM} que aparece como productos de las matrices P_s .

Este último cálculo es complicado normalmente ya que aparece el producto de S matrices en las que intervienen los gradientes de las funciones g_s y la inversa de la matriz \mathbf{I}_{oc} .

Un caso particular de algoritmo ECM es el algoritmo PECM que consiste en elegir las funciones g_s correspondiendo a una partición del espacio paramétrico,

este algoritmo hace muy simple el cálculo de la matrix DM^{CM} . Así, si θ está dividido en S subvectores, $\theta = (\vartheta_1, \dots, \vartheta_S)$, donde cada ϑ_s es de dimensión r_s , las funciones g_s están definidas por $g_s = (\vartheta_1, \dots, \vartheta_{s-1}, \vartheta_{s+1}, \dots, \vartheta_S)$. Sea Υ la matriz diagonal por bloques de \mathbf{I}_{oc} con S bloques correspondiendo a los S subvectores de θ . Sea Γ la correspondiente matriz por bloques triangular inferior, es decir, $\mathbf{I}_{oc} = \Upsilon + \Gamma + \Gamma'$. Entonces en este caso (van Dyk el al. (1994), p. 8)

$$DM^{CM} = -\Gamma(\Upsilon + \Gamma')^{-1}. \quad (1.34)$$

La regla de parada (precisión) para los algoritmos suplementados debe de ser menos exigente que para los algoritmos sin suplementar, ya que para el cálculo de la matrix $DM^{algoritmo}$ se utiliza una aproximación numérica lo cual es menos exacto que la evaluación de la función directamente. Por esto la regla de parada en estos casos se suele tomar como la raíz cuadrada de la misma cuando el algoritmo no es suplementado.

La matriz de varianzas-covarianzas, V , que se obtiene teóricamente debe de ser simétrica. Esta propiedad nos da un método para diagnosticar si el modelo es válido, si hay errores de programación o de precisión numérica, etc. De hecho esto nos hizo modificar nuestros desarrollos mejorando los resultados obtenidos. Incluso, puede servir para chequear si el punto estacionario obtenido es el estimador de máxima verosimilitud o no lo es, ya que muchos algoritmos erróneos convergen.

Capítulo 2

Elementos sobre campos espaciales con parámetro discreto

2.1 Introducción

En este capítulo se reúnen una serie de modelos y técnicas estadísticas concernientes principalmente a procesos estocásticos estacionarios sobre \mathbb{Z}^2 . Los resultados que presentamos aparecen en una serie de artículos destacando el pionero de Whittle (1954) y el de Besag (1974).

Como se sabe, el estudio de procesos en dos dimensiones no es una mera extensión de los resultados obtenidos en una. Básicamente ello está motivado por el hecho de que \mathbb{Z}^2 no está naturalmente ordenado y los polinomios en varias variables no poseen, en general, las buenas propiedades de los polinomios en una variable. Introducimos los procesos ARMA espaciales, en términos de operadores de retardos, e introducimos la noción de causalidad en el espacio como una extensión de la causalidad en una dimensión, con referencia a la adopción de algún orden (en particular, lexicográfico horizontal). Presentamos una extensión de los modelos Markovianos, para la cual, en una primera etapa, es necesario definir el concepto de “vecindad” y, posteriormente, dados un sistema de vecinos, obtener la forma general de una función llamada negpotencial que nos permitirá, en cierta medida, calcular la probabilidad conjunta de cualquier observación. Finalmente, mostramos diversos modelos y esquemas de los cuales destaca, por su amplio uso, el de “vecino más próximo”. Una subclase importante la constituyen los procesos separables; se analiza, en particular, el modelo $\text{AR}(1) \times \text{AR}(1)$ del cual se señalan algunas propiedades

fundamentales. Finalizamos mostrando algunos métodos de estimación de los parámetros.

2.2 Procesos ARMA sobre \mathbb{Z}^2

Para la presentación de los procesos ARMA en el plano \mathbb{Z}^2 introducimos dos operadores de retardos B_1 y B_2 definidos sobre un proceso bidimensional como sigue:

$$B_1 X_{ij} = X_{i-1,j} \qquad B_2 X_{ij} = X_{i,j-1}.$$

Sean P y N dos funciones homomorfas en un entorno del conjunto T^2 de \mathbb{C}^2 , siendo

$$T = \{z \in \mathbb{C} : |z| = 1\}.$$

Se supondrá además que:

$$P(z_1, z_2) \neq 0 \qquad \text{si} \qquad |z_1| = |z_2| = 1, \qquad (2.1)$$

por lo que $P^{-1}N$ existe sobre T^2 y admite un desarrollo de Laurent en un entorno de T^2 :

$$P^{-1}N(z_1, z_2) = \sum_{\mathbb{Z}^2} c_{kl} z_1^k z_2^l. \qquad (2.2)$$

Sea $\epsilon = \{\epsilon_t : t \in \mathbb{Z}^2\}$ un ruido blanco, es decir

$$E(\epsilon_t) = 0, \quad \text{Var}(\epsilon_t) = \sigma_\epsilon^2$$

ϵ_t es ortogonal a ϵ_s si $s \neq t$.

Definición 2.2.1

Un proceso espacial $X = \{X_t : t \in \mathbb{Z}^2\}$ es un proceso estacionario de segundo orden si verifica:

1. X_t es de cuadrado integrable para todo t .
2. $E[X_t] = \mu$ para todo t .
3. $\text{Cov}[X_t, X_s] = r(t-s)$.

Definición 2.2.2

Se dirá que un proceso estacionario X es un proceso ARMA generalizado si existen dos funciones P y N homomorfas sobre T^2 , con P verificando (2.1), y un ruido blanco $\epsilon = \{\epsilon_t : t \in \mathbb{Z}^2\}$ tales que:

$$P(B_1, B_2)X_{ij} = N(B_1, B_2)\epsilon_{ij} \quad (2.3)$$

Debido a la expresión (2.2), el proceso X adopta también la expresión:

$$X_{ij} = \sum_{\mathbb{Z}^2} c_{uv} \epsilon_{i-u, j-v}.$$

Se supondrá además:

$$N(z_1, z_2) \neq 0 \quad \text{si} \quad |z_1| = |z_2| = 1, \quad (2.4)$$

que es una condición similar a la de invertibilidad para series temporales.

La densidad espectral de un proceso ARMA generalizado viene dada por

$$f(\lambda_1, \lambda_2) = \frac{\sigma^2 |N(e^{i\lambda_1}, e^{i\lambda_2})|^2}{4\pi^2 |P(e^{i\lambda_1}, e^{i\lambda_2})|^2}, \quad \lambda_1, \lambda_2 \in [0, 2\pi). \quad (2.5)$$

Interesa, desde el punto de vista de la inferencia, considerar procesos que puedan representarse mediante una parametrización finita. Se dirá que P o N son polinomios si los desarrollos de P o N son finitos, es decir

$$\begin{aligned} P(z_1, z_2) &= \sum_R a_{kl} z_1^k z_2^l && R \subset \mathbb{Z}^2, R \text{ finito} \\ N(z_1, z_2) &= \sum_S b_{kl} z_1^k z_2^l && S \subset \mathbb{Z}^2, S \text{ finito} \end{aligned} \quad (2.6)$$

Se dirá que P es unilateral si $R \subseteq (\mathbb{Z}^2)^+$; en otro caso se dirá que P es bilateral. También se supondrá la condición de normalización

$$a_{00} = b_{00} = 1.$$

Definición 2.2.3

Un proceso estocástico $X = \{X_{st}, (s, t) \in \mathbb{Z}^2\}$ es un proceso ARMA si existen dos polinomios P y N y un ruido blanco ϵ tales que:

$$P(B_1, B_2)X_{ij} = N(B_1, B_2)\epsilon_{ij}, \quad (2.7)$$

con P y N verificando las condiciones (2.1) y (2.4).

La densidad espectral del proceso X también vendrá dada, obviamente, por la expresión (2.5). Si $P \equiv 1$ diremos que el proceso es de medias móviles (MA). Si $N \equiv 1$ diremos que el proceso es autorregresivo (AR). Los modelos AR espaciales fueron estudiados por primera vez por Whittle (1954) y han sido objeto de numerosos trabajos, entre los que caben destacar los de Ord (1975), Besag (1977) y (1981), Martin (1990), ... Los procesos MA espaciales han sido estudiados, en particular, por Haining (1978). Otros autores como Guyon (1990), Huang et al. (1992), Vecchia (1988), ... se refieren a procesos ARMA espaciales en general.

Como en el caso de series temporales, el interés de la modelización ARMA está en el hecho de que dada una realización de un proceso con densidad espectral racional, esta puede ser "representada" mediante un modelo ARMA de dimensiones paramétricas relativamente factibles (es decir, R y S "pequeños").

Definición 2.2.4

Sea X un proceso espacial; diremos que el proceso ϵ es la innovación de X si $\epsilon_{ij} = X_{ij} - P_L[X_{ij} | \text{esp}\{X_{i-1,j}\}]$, siendo $\text{esp}\{X_{ij}\}$ el espacio generado por las variables X_{kl} con $l < j$ o $l = j$ y $k \leq i$, y P_L la proyección de la variable sobre dicho espacio.

Definición 2.2.5

Diremos que el proceso definido por la ecuación (2.7) es una representación ARMA causal de X si ϵ es la innovación de X .

Es claro que para un proceso MA espacial una condición necesaria y suficiente para causalidad es que el polinomio N sea unilateral. Una condición necesaria y suficiente para la causalidad en un proceso AR espacial viene dada por el siguiente teorema.

Teorema 2.2.6 (Guyon (1990), pp. 15-16)

La representación de la definición anterior será AR causal si y solo si se satisfacen las dos condiciones siguientes:

1. P es un polinomio unilateral.
2. Se verifica una de las siguientes propiedades:
 - (a) • $P(z_1, 0) \neq 0$ si $|z_1| \leq 1$.

- $P(e^{i\lambda_1}, z_2) \neq 0$, si $|z_2| \leq 1$, y para todo $\lambda_1 \in \mathbb{R}$.

$$(b) \int_{T^2} \log |P(e^{i\lambda_1}, e^{i\lambda_2})| d\lambda_1 d\lambda_2 = 0$$

Una condición suficiente que asegure unilateralidad es en este caso:

$$P(z_1, z_2) = \sum_{(k,l) \geq 0} a_{kl} z_1^k z_2^l, \quad a_{00} = 1. \quad (2.8)$$

$$\sum_{(k,l) > 0} |a_{kl}| < a_{00} = 1 \quad (2.9)$$

Si la estructura autorregresiva no es causal, se dirá que la representación es bilateral.

Observación 2.2.7 *Contrariamente a como ocurre en \mathbb{Z} , un proceso sobre \mathbb{Z}^2 autoregresivo, no admite en general una escritura autorregresiva causal finita.*

Las siguientes representaciones (sobre \mathbb{Z}^2 y \mathbb{Z} respectivamente) son bilaterales:

$$X_{ij} = \alpha(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}) + \epsilon_{ij}, \quad |\alpha| < 1/4$$

$$X_t = \alpha(X_{t-1} + X_{t+1}) + \epsilon_t, \quad |\alpha| < 1/2$$

la primera no admite una representación AR unilateral equivalente.

Observación 2.2.8 *Contrariamente a la situación en \mathbb{Z} , un proceso sobre \mathbb{Z}^2 con covarianza nula fuera de un subconjunto finito de \mathbb{Z}^2 no admite en general, una representación MA.*

Análogamente, para un proceso ARMA espacial la causalidad se caracteriza por el siguiente resultado:

Teorema 2.2.9 (Guyon (1990), pp. 19-20)

Sea un proceso ARMA definido por la ecuación $P(B_1, B_2)X_{ij} = N(B_1, B_2)\epsilon_{ij}$, siendo P y N tales que verifican las condiciones del teorema 2.2.6. Entonces ϵ es la innovación del proceso X .

2.3 Modelos Markovianos

Los procesos de Markov son muy útiles en una dimensión y se han realizado numerosos intentos para definir procesos con análogas propiedades en espacios arbitrarios, en particular en el plano. Todos ellos parecen susceptibles de crítica, ya que la propiedad de orden en la línea (un punto divide la línea en pasado y futuro) es fundamental en la teoría de los procesos de Markov. En lugar de pasado y futuro necesitamos una regla que nos permita, dado un par de puntos, decidir si son o no “vecinos”. Para puntos en una malla a menudo estas reglas son simples. En puntos del plano usualmente se consideran vecinos puntos que distan menos de una cantidad dada. Dado el conjunto de vecinos de cada subconjunto, se puede definir el proceso de Markov de la siguiente forma:

Definición 2.3.1

Dado un conjunto A arbitrario y siendo $E(A)$ el conjunto de vecinos de A , diremos que un proceso es de Markov si la probabilidad condicionada del proceso en los puntos A al resto de las variables del proceso, sólo depende del proceso en los puntos $E(A) \setminus A$.

Obviamente, que el proceso sea o no de Markov depende de la elección del conjunto $E(A)$ de vecinos. Claramente, para un proceso, este siempre va a ser de Markov si definimos los vecinos de cada punto como todo el conjunto de nodos donde se realiza el proceso. Esto indica, que dado un proceso, siempre podremos definir los vecinos de un punto (eligiendo un conjunto más o menos amplio) de forma que el proceso sea de Markov. En el caso unidimensional, se puede ver que la definición anterior nos indica que un proceso es de Markov si y sólo si el conjunto de vecinos del punto t se puede expresar de la forma $E(t) = \{t + 1, t - 1\}$. Determinados conjuntos de vecinos dan lugar a esquemas de interés, como los modelos de “vecino más próximo”, que parece una primera generalización al caso de dos dimensiones de la propiedad de Markov, en la cual los vecinos del punto (i, j) son $\{(i + 1, j), (i - 1, j), (i, j - 1), (i, j + 1)\}$.

Sin embargo, el problema de modelizar un proceso según el esquema de “vecino más próximo” no es fácil. Whittle (1963) y Bartlett (1955) definen este problema de dos formas diferentes. Para Whittle, dada una malla, se dice que se está en un esquema de vecino más próximo si la probabilidad conjunta

se puede descomponer en un producto de la forma:

$$\prod_{ij} Q(x_{ij}; x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}).$$

Bartlett define el esquema de vecino más próximo, como la situación en que la distribución de X_{ij} condicionada a los restantes nodos sólo depende de las variables $X_{i-1,j}$, $X_{i+1,j}$, $X_{i,j-1}$ y $X_{i,j+1}$. Sin embargo, a pesar de ser más intuitiva, esta formulación presenta algunas desventajas:

1. No hay un método obvio para deducir la estructura de probabilidad conjunta.
2. La estructura de probabilidad condicionada está sujeta a condiciones de consistencia altamente restrictivas.
3. La especificación natural de un proceso de equilibrio en mecánica estadística se hace en términos de su probabilidad conjunta más bien que en términos de la distribución condicionada de las variables.

Con objeto de ilustrar que el enfoque condicional de Bartlett es equivalente al simultáneo de Whittle cuando conocemos una probabilidad conjunta en un punto, consideramos observaciones discretas en los nodos y trabajaremos con funciones masa de probabilidad. La extensión al caso de variables aleatorias continuas y funciones de densidad es obvia. Los resultados que mostramos son válidos para esquemas más generales que el de vecino más próximo, pudiendo incluso estar los nodos irregularmente espaciados.

Hipótesis 2.3.2 (Condición de positividad)

Si x_1, \dots, x_n pueden ocurrir individualmente en los nodos $1, \dots, n$ respectivamente, entonces pueden ocurrir juntos, es decir:

$$P(x_i) > 0, \forall i \Rightarrow P(x_1, \dots, x_n) > 0.$$

Esta condición es satisfecha en la práctica normalmente. La supondremos en el resto del capítulo.

Definimos el espacio muestral Ω como el conjunto de todas las posibles realizaciones del sistema, $\mathbf{x} = (x_1, \dots, x_n)$. Es decir, $\Omega = \{\mathbf{x} : P(\mathbf{x}) > 0\}$.

Teorema 2.3.3

Sean \mathbf{x} e $\mathbf{y} \in \Omega$. Entonces:

$$\frac{P(\mathbf{x})}{P(\mathbf{y})} = \prod_{i=1}^n \frac{P(x_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)}{P(y_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)} \quad (2.10)$$

A la vista de este resultado, y debido a la gran cantidad de valores que se podrían presentar en los nodos, podemos deducir que habrá gran cantidad de factorizaciones como la del teorema anterior (una para cada \mathbf{y}). También es claro que si se conoce la probabilidad de una realización \mathbf{y} , y las probabilidades condicionadas, se puede calcular cualquier probabilidad no condicionada.

Vamos a definir como tomar el conjunto de vecinos de un punto para que el proceso sea de Markov

Definición 2.3.4

Se dice que el nodo j es un vecino de i ($j \neq i$) si y sólo si la forma funcional $P(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ depende de x_j .

De la definición de Bartlett, que es la que adoptaremos, ya que las probabilidades conjuntas las podremos calcular a partir del teorema anterior, el esquema del vecino más próximo nos conduce a decir que los vecinos del nodo (i, j) son los nodos $(i-1, j)$, $(i+1, j)$, $(i, j-1)$ y $(i, j+1)$ (también es llamado esquema de primer orden).

Cualquier sistema de n nodos, cada uno con sus vecinos especificados, claramente genera una clase de esquemas estocásticos. A cada miembro de esta clase se le llama campo de Markov.

Definición 2.3.5 (clique o conjunto de vecinos simultáneos)

Se llama clique a cualquier conjunto de nodos que contenga un solo nodo o tal que cada nodo que contenga sea vecino de los restantes nodos.

Así en el esquema de vecino más próximo, los cliques son conjuntos de la forma $\{(i, j)\}$, $\{(i, j), (i, j-1)\}$ y $\{(i, j), (i-1, j)\}$.

Sin pérdida de generalidad, podemos asumir que el 0 se puede observar en cada nodo, con lo que $\mathbf{0} \in \Omega$. Definimos

$$Q(\mathbf{z}) \equiv \log\{P(\mathbf{z})/P(\mathbf{0})\}, \quad \mathbf{z} \in \Omega. \quad (2.11)$$

La función Q se llama función negpotencial y es claro que su conocimiento implica el conocimiento de la función P , ya que

$$P(\mathbf{z}) = \frac{\exp(Q(\mathbf{z}))}{\sum_{\mathbf{y} \in \Omega} \exp(Q(\mathbf{y}))}$$

Dado $\mathbf{x} \in \Omega$, vamos a notar \mathbf{x}_i a la realización

$$\mathbf{x}_i = (x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$$

Hammersley y Clifford responden a la siguiente pregunta: ¿Cual es la forma más general de la función Q dados los entornos de cada punto, en orden a que proporcione una estructura de probabilidad válida para el sistema?. Debido a que

$$\exp\{Q(\mathbf{x}) - Q(\mathbf{x}_i)\} = \frac{P(\mathbf{x})}{P(\mathbf{x}_i)} = \frac{P(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}{P(0 | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}, \quad (2.12)$$

este problema nos da la solución de la forma más general que puede tomar una probabilidad condicionada en cada nodo.

Teorema 2.3.6 (Hammersley-Clifford, Besag (1974), p. 197)

Dada una función $P(\mathbf{x})$ verificando las restricciones anteriores, existe una expansión de $Q(\mathbf{x})$ única sobre Ω de la forma

$$\begin{aligned} Q(\mathbf{x}) = & \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i < j \leq n} x_i x_j G_{i,j}(x_i, x_j) + \\ & + \sum_{1 \leq i < j < k \leq n} x_i x_j x_k G_{i,j,k}(x_i, x_j, x_k) + \dots + \\ & + x_1 x_2 \dots x_n G_{1,\dots,n}(x_1, \dots, x_n), \end{aligned} \quad (2.13)$$

donde la función $G_{i,j,\dots,s}$ es no nula si y sólo si los nodos i, j, \dots, s forman un clique.

Por lo tanto, dados los entornos de cada punto, podemos dar la forma más general de la función Q y, por tanto, de las distribuciones condicionadas.

Corolario 2.3.7

Para un campo de Markov,

$$P(X_i = x_i, X_j = x_j, \dots, X_s = x_s | \text{ todos los demás })$$

depende de x_i, x_j, \dots, x_s y los valores en los nodos vecinos a i, j, \dots, s .

Corolario 2.3.8 (Cressie (1991), pp. 418-419)

Supongamos que Ω es numerable (o Lebesgue-medible) en \mathbb{R}^d . Sean G funciones como en el teorema de Hammersley-Clifford. Entonces la función Q resultante nos conduce a una única probabilidad proporcional a $\exp\{Q(\mathbf{x})\}$ dada la condición de sumabilidad

$$\sum_{z \in \Omega} \exp\{Q(z)\} < \infty \quad \left(\text{o } \int_{\Omega} \exp\{Q(z)\} dz < \infty \right)$$

2.3.1 Algunos esquemas espaciales

Automodelos

Se denominan automodelos aquéllos que tienen una función Q de la forma

$$Q(\mathbf{x}) = \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i < j \leq n} x_i x_j G_{i,j}(x_i, x_j)$$

Esta definición es consecuencia de las siguientes dos hipótesis:

[H1] Los cliques tienen a lo sumo dos nodos.

[H2] La distribución de probabilidad condicionada pertenece a la familia exponencial, esto es

$$\ln P(x_i \mid \text{ todos los demás }) = A_i(\cdot) B_i(x_i) + C_i(x_i) + D_i(\cdot),$$

donde B_i y C_i son de cierta forma especificada y A_i y D_i dependen sólo de los vecinos de x_i .

Proposición 2.3.9 (Besag (1974), pp. 202-203)

Como consecuencia de las hipótesis anteriores se puede probar que

$$A_i(\cdot) = \alpha_i + \sum_{j=1}^n B_j(x_j) \beta_{i,j}$$

donde $\beta_{i,j} = \beta_{j,i}$ y $\beta_{i,j} = 0$ si los nodos i y j no pertenecan a un clique.

Por lo tanto, de la proposición anterior se deduce que la función $G_{i,j}$ adopta la expresión

$$G_{i,j}(x_i, x_j) \equiv \beta_{i,j} H_i(x_i) H_j(x_j),$$

donde $x_i H_i(x_i) = B_i(x_i) - B_i(0)$.

Modelos en malla

El modelo más sencillo es el esquema de vecino más próximo o esquema de primer orden, cuyos vecinos se han especificado anteriormente. Así, la función Q adopta la expresión

$$Q(\mathbf{x}) = \sum x_{i,j} \phi_{i,j}(x_{i,j}) + \sum x_{i,j} x_{i+1,j} \psi_{1,i,j}(x_{i,j}, x_{i+1,j}) + \\ + \sum x_{i,j} x_{i,j+1} \psi_{2,i,j}(x_{i,j}, x_{i,j+1}),$$

donde $\{\phi_{i,j}\}$, $\{\psi_{1,i,j}\}$ y $\{\psi_{2,i,j}\}$ son conjuntos de funciones arbitrarias. Si el esquema es homogéneo las anteriores funciones serán constantes y si es isotrópico, $\psi_1 = \psi_2$.

Este esquema es fácilmente extensible para producir esquemas de ordenes más altos. Así se define el esquema de segundo orden como el que tiene como vecinos del nodo (i, j) los nodos $(i+1, j-1)$, $(i+1, j)$, $(i+1, j+1)$, $(i, j-1)$, $(i, j+1)$, $(i-1, j-1)$, $(i-1, j)$ y $(i-1, j+1)$. Al esquema de tercer orden se le añaden los nodos $(i-2, j)$, $(i+2, j)$, $(i, j-2)$ y $(i, j+2)$. Puede verse entonces que, por ejemplo, en el esquema de segundo orden los cliques vienen dados por $\{(i, j)\}$, $\{(i, j), (i+1, j)\}$, $\{(i, j), (i, j+1)\}$, $\{(i, j), (i+1, j+1)\}$, $\{(i, j), (i+1, j-1)\}$, $\{(i, j), (i+1, j), (i, j+1)\}$, $\{(i, j), (i+1, j), (i+1, j+1)\}$, $\{(i, j), (i+1, j), (i+1, j-1)\}$, $\{(i, j), (i, j+1), (i+1, j+1)\}$ y $\{(i, j), (i+1, j), (i, j+1), (i+1, j+1)\}$.

2.4 Procesos separables

Una subclase importante de los procesos espaciales la constituyen los procesos separables. Este tipo de procesos se puede definir, cuando el proceso es estacionario, por la propiedad de que la correlación en un retardo (k, l) puede calcularse como el producto de la correlación en $(k, 0)$ por la correlación en $(0, l)$. Esto indica que la estructura de correlación del proceso se puede ver como la confluencia de las correlaciones de dos procesos unidimensionales a lo largo de las filas y las columnas. Se puede ver que si Z_1 y Z_2 son dos procesos que tienen la estructura de correlación del proceso separable X bidimensional, cada uno en una de las direcciones, con representación ARMA dada por los polinomios P_i y N_i , entonces X tiene una representación ARMA

$$P_1(B_1)P_2(B_2)X_{ij} = N_1(B_1)N_2(B_2)\epsilon_{ij}, \quad (2.14)$$

donde B_1 y B_2 son los operadores de retardos.

Teorema 2.4.1 (Martin (1979), p. 211)

Si X es un proceso dado por (2.14) entonces la matriz de correlación de X es el producto de Kronecker de las matrices de correlación de Z_1 y Z_2 .

La clase de los procesos separables es claramente una subclase propia de los procesos estacionarios; por ejemplo, estos procesos tienen la propiedad de reflexión simétrica, esto es, la correlación (k,l) es la misma que la correlación $(k,-l)$. Esta propiedad raramente es una desventaja, ya que la mayoría de los modelos postulados en aplicaciones poseen esta propiedad. Más aún, esta subclase ha resultado ser muy útil en la práctica. El proceso más simple, extensión natural de los procesos de Markov unidimensionales al plano, y que puede ser la aproximación del proceso en múltiples aplicaciones, es el proceso $AR(1) \times AR(1)$ o proceso doblemente geométrico (ver Martin (1979)). Así, este proceso viene dado por la confluencia de dos procesos $AR(1)$ en las dos direcciones, viniendo su expresión dada por

$$X_{ij} = \lambda X_{i-1,j} + \beta X_{i,j-1} - \lambda\beta X_{i-1,j-1} + \epsilon_{ij}.$$

Aunque es claramente irreal como un modelo para datos con una estructura de correlación continuo-espacial isotrópica cuyo decrecimiento sea "suave" (ver Besag (1981)), ha sido de gran uso (la correlación decae exponencialmente en este modelo).

Algunas propiedades de este proceso son las siguientes:

1. Su función generatriz de covarianzas $\Gamma_x(\omega_1, \omega_2)$ es proporcional a

$$\{(1 - \lambda\omega_1)(1 - \lambda\omega_1^{-1})(1 - \beta\omega_2)(1 - \beta\omega_2^{-1})\}^{-1}.$$

2. Las covarianzas, por tanto, vienen dadas por:

$$C_{(i,j),(k,l)}^{XX} = \frac{\lambda^{|i-k|} \beta^{|j-l|} \sigma_\eta^2}{(1 - \lambda^2)(1 - \beta^2)}.$$

3. Su correlación, por

$$R_{(i,j),(k,l)}^{XX} = \lambda^{|i-k|} \beta^{|j-l|}.$$

4. Las autocorrelaciones parciales $\rho_{h,r,k,s}$ son cero para $0 < r < h$ y $0 < s < k$.

5. Las autocorrelaciones inversas están dadas por $p_{1,0} = -\alpha$, $p_{0,1} = -\gamma$, $p_{1,1} = \alpha\gamma$ y $p_{hk} = 0$, para $|h|, |k| > 1$, donde $\alpha = \lambda/(1 + \lambda^2)$, $\gamma = \beta/(1 + \beta^2)$.
6. La representación condicional es

$$E[x_{ij}|\cdot] = \lambda(x_{i-1,j} + x_{i+1,j}) + \beta(x_{i,j-1} + x_{i,j+1}) - \lambda\beta(x_{i-1,j-1} + x_{i-1,j+1} + x_{i+1,j-1} + x_{i+1,j+1}).$$

Otras generalizaciones de este tipo de procesos se presentan en el próximo capítulo.

2.5 Métodos de estimación

En esta sección suponemos que tenemos un proceso X estacionario, centrado, dependiendo de un parámetro θ , $\theta \in \Theta$, siendo Θ un compacto de \mathbb{R}^n con interior no vacío.

Vamos a estudiar dos aproximaciones para la estimación de dicho parámetro. La primera consiste en el cálculo de la log-verosimilitud o de una aproximación de la misma, y calcular el valor θ que la haga máxima. La segunda consiste en utilizar la forma markoviana de X y utilizar el método "coding" de Besag (1974).

A partir de ahora supondremos además que el proceso es gaussiano.

Notación:

$|N|$ denota el número de nodos.

Γ_N es una matriz $|N| \times |N|$ que contiene las covarianzas de proceso.

2.5.1 Estimación máximo-verosímil

Bajo las hipótesis anteriores la log-verosimilitud de X viene dada por

$$L = -\frac{|N|}{2} \log 2\pi - \frac{1}{2} \log |\Gamma_N| - \frac{1}{2} X' \Gamma_N^{-1} X.$$

La matriz Γ_N se puede factorizar de la forma

$$\Gamma_N = kV^{-1},$$

donde V es la matriz de los coeficientes de las esperanzas condicionadas de X_i al resto. k se interpreta como:

- k , factor de variabilidad, en los modelos con espectro racional ($f = k \frac{G}{H}$, con las condiciones de normalización $g_0 = h_0 = 1$).
- $k = \sigma_\epsilon^2$, varianza de la innovación, en modelos ARMA causales.
- $k = \sigma_\epsilon^2$, varianza del ruido generador, en modelos ARMA no causales.

Maximizar L equivale a minimizar:

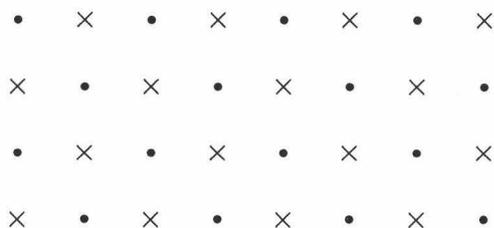
$$L' = |N| \log k - \log |V| + \frac{1}{k} X' V X \quad (2.15)$$

Ahora hay que minimizar la función anterior para obtener la estimación de los parámetros. Si X es gaussiano (como supusimos) el estimador obtenido es el de máxima verosimilitud. Si X no fuese gaussiano el estimador obtenido por este método se denomina estimador de máxima pseudoverosimilitud.

2.5.2 Métodos coding

Estos métodos consisten básicamente en simplificar un poco la función de verosimilitud para posteriormente aplicar EMV sobre dicha función. La simplificación consiste en tomar nodos que sean independientes entre sí al condicionarlos al resto de los nodos. Al ser independientes, la verosimilitud queda como un producto de cada nodo condicionado al resto de la malla (aunque en realidad sólo aparece condicionado a sus vecinos).

Así, para un esquema de primer orden, una partición sería



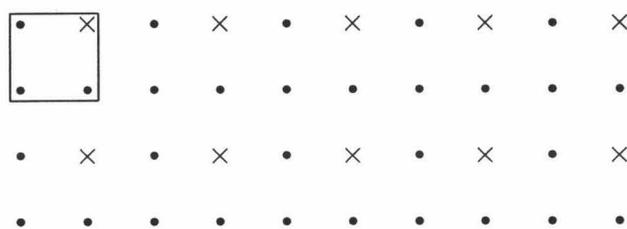
Las observaciones en los nodos \times son independientes entre sí (en los nodos punto también) dadas el resto. Por esto la verosimilitud se podrá expresar de la forma

$$\prod P(x_{ij} \mid x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1})$$

donde los x_{ij} responden a observaciones situadas en los nodos \times . Ahora simplemente estimamos los parámetros maximizando la función anterior.

Hay que hacer notar que si tomamos las observaciones en los nodos punto y las condicionamos a las demás, obtenemos un desarrollo análogo y, por lo tanto, otros estimadores para el mismo modelo. Estos dos procedimientos, dependiendo de qué nodos tomemos, son altamente dependientes (son los mismos datos) pero es razonable en la práctica calcularlos y combinar las estimaciones de algún modo.

En el esquema de segundo orden uno de los posibles esquemas *coding* será

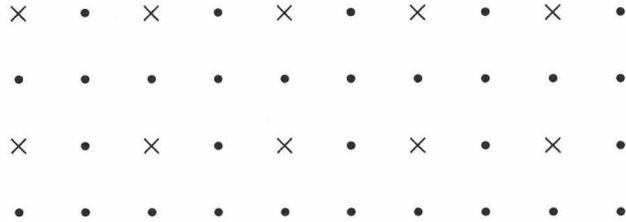


Como puede verse, este esquema genera cuatro estimaciones distintas (intercambiando \times por punto dentro del cuadrado) que pueden ser combinadas apropiadamente.

Estos métodos pueden conducir a formular tests de cocientes de verosimilitudes, para examinar la bondad de ajuste de tales esquemas, con ciertas salvedades:

1. La hipótesis nula debe ser propuesta a priori (como ocurre usualmente), es decir, se debe tener alguna intuición previa sobre qué modelo se puede presentar.
2. Las dos verosimilitudes del cociente deben ser comparables; por ejemplo, si contrastamos un modelo de primer orden frente a un modelo de segundo orden, las dos representaciones anteriores son comparables. Sin embargo la primera no es comparable con





3. Hay más de un test disponible (trasladando \times) y hay que considerarlos colectivamente.

2.5.3 Estimación de procesos ARMA

Cuando estamos interesados en estimar un proceso ARMA, nuestro problema se divide en dos etapas. En la primera etapa, habrá que estimar o calcular de alguna forma los grados de los polinomios de retardos, y en una segunda etapa, estimar los coeficientes de los mismos. En este apartado vamos a presentar algunos resultados relativos a la primera etapa. La estimación de los parámetros se realizará por uno de los métodos anteriormente expuestos.

Lema 2.5.1 (Huang et al. (1992), p. 519)

Bajo la condición 2(a) del teorema 2.2.6, y siendo $N(z) = \sum_{j \in G(q)} b_j z^j$, existe $\{\phi_t : t \geq 0\}$ tal que, para $z \in T^2$,

$$\frac{\sum_{k \in G(p)} a_k z^k}{\sum_{j \in G(q)} b_j z^j} = \sum_{t \geq 0} \phi_t z^t \quad (\phi_0 = 1) \quad (2.16)$$

$$\sum_{j \in G(q)} b_j \phi_{t-j} = \begin{cases} a_t & t \in G(p) \\ 0 & t \notin G(p) \end{cases} \quad (2.17)$$

Como resultado de 2.16 definimos

$$Y_t = \sum_{k \geq 0} \phi_k \epsilon_{t-k} \quad (\phi_0 = 1), \quad (2.18)$$

donde $\{\epsilon_t\}$ es el proceso de ruido blanco espacial de (2.7). A la vista de (2.17), el proceso Y satisface la ecuación

$$\sum_{j \in G(q)} b_j Y_{t-j} = \sum_{k \in G(p)} a_k \epsilon_{t-k}. \quad (2.19)$$

Definamos $G'(p) = G(p) \cap S$, donde $S = \{t : t > 0\} = \{(t_1, t_2) : 1 \leq t_1, \text{ o } t_1 = 0 \text{ y } t_2 > 0\}$. Definamos también:

$$\sigma_{pq}^2 = \|Y_0 - E[Y_0 | \epsilon_{-t}, t \in G'(p); Y_{-j}, j \in G'(q)]\|^2, \quad (2.20)$$

$$\mu_{pq}^2 = \frac{\|A\|^2}{\sigma^2}, \quad (2.21)$$

donde $\sigma^2 = E[\epsilon_t]$, y

$$A = E[Y_0 | \epsilon_{-t} : t \in G'(p), t > 0] - E[E[Y_0 | \epsilon_{-t} : t \in G'(p), t > 0] | E[Y_{-j} | \epsilon_{-t} : t \in G'(p), t > 0] : j \in G'(q)].$$

Condición de identificabilidad

Dadas dos funciones complejo-valuadas $\sum_{k \in G(p)} a_k z^k$ y $\sum_{j \in G(q)} b_j z^j$, si existen dos funciones complejo-valuadas $\sum_{k \in G(p')} \alpha_k z^k$ y $\sum_{j \in G(q')} \beta_j z^j$ con $\sum_{j \in G(q')} |\beta_j|^2 > 0$ tales que

$$\frac{\sum_{k \in G(p)} a_k z^k}{\sum_{j \in G(q)} b_j z^j} = \frac{\sum_{j \in G(q')} \beta_j z^j}{\sum_{k \in G(p')} \alpha_k z^k} \quad (z \in T^2),$$

entonces

$$G(p) \subset G(p') \quad \text{y} \quad G(q) \subset G(q').$$

Teorema 2.5.2 *Bajo la condición de identificabilidad anterior,*

1. $\mu_{p'q'}^2 = 0$ si y sólo si $G(q) \subset G(q')$ y $G(p) \subset G(p')$.

2. $G(q) \not\subset G(q') \Rightarrow \min_{b_j : \sum_{j \in G(q')} |b_j|^2 \geq 1} \sum_{t > 0, t \notin G'(p)} \left| \sum_{j \in G(q')} b_j \phi_{t-j} \right|^2 > 0.$

Este teorema nos muestra un método para identificar los órdenes apropiados de los polinomios de retardos.

Capítulo 3

Estimación en cierto tipo de procesos espaciales con datos faltantes

3.1 Introducción

Uno de los problemas principales que suele presentarse en el análisis de datos espaciales (sobre una malla) es que en la mayoría de los casos no se dispone de una información exhaustiva de todos los nodos sino que en algunos puede ocurrir que falten observaciones. Esto puede ocurrir, bien porque haya imposibilidad o inconveniencia (método muy costoso, accidentes geográficos, etc.) en el procedimiento y sólo se elijan algunos nodos, bien porque las observaciones de que disponemos se hayan tomado con anterioridad y posteriormente se le haya querido ajustar una malla, de tal forma que las observaciones prácticamente coincidan con algún nodo, quedando claramente nodos sin observación. Por ello, debemos de desarrollar una metodología que permita construir modelos en situaciones en que haya *observaciones faltantes*. Dempster et al. (1977) desarrollaron una metodología denominada algoritmo EM que se puede aplicar a situaciones con observaciones faltantes para variables aleatorias. Shumway et al. (1982) aplicaron esta metodología a series temporales.

En dicha generalización se considera un proceso AR(1) multivariante como modelo para la serie debido a que cualquier proceso AR(p) puede ser transformado en un proceso AR(1) de dimensión mayor. Nosotros vamos a proponer modelizar el proceso espacial por un proceso separable $AR(1) \times AR(1)$. Es-

tos modelos han sido estudiados ampliamente por Martin (1979) y (1990). No obstante, proponemos unas generalizaciones a modelos $AR(1) \times AR(2)$, $AR(2) \times AR(2)$, $MA(1) \times AR(1)$, $MA(1) \times MA(1)$, y modelos más complicados.

Además de calcular los parámetros del modelo, vamos a ser capaces de estimar las observaciones faltantes mediante un proceso de alisamiento, que básicamente consiste en calcular las esperanzas condicionadas, suponiendo el modelo, al resto de las observaciones no faltantes. Más aún, una vez que tengamos estimado nuestro modelo, dispondremos de herramientas para estimar el valor que puede tomar el proceso espacial, fuera de la malla finita de la que partimos.

3.2 Modelo

Consideramos que el proceso de interés $\{X_{ij}\}_{(i,j) \in \mathbb{Z}^2}$ tiene una estructura doblemente geométrica (Martin, 1979):

$$X_{ij} = \lambda X_{i-1,j} + \beta X_{i,j-1} - \lambda\beta X_{i-1,j-1} + \eta_{ij}$$

donde $\eta_{ij} \sim N(0, \sigma_\eta^2)$ son independientes y los coeficientes λ y β son constantes con módulo menor que 1. Así la covarianza entre las variables X_{ij} y X_{kl} viene dada por

$$C_{(i,j),(k,l)}^{XX} = \frac{\lambda^{|i-k|} \beta^{|j-l|} \sigma_\eta^2}{(1-\lambda^2)(1-\beta^2)},$$

y su correlación por

$$R_{(i,j),(k,l)}^{XX} = \lambda^{|i-k|} \beta^{|j-l|}.$$

Consideremos el problema de estimación y alisamiento (a partir de datos disponibles) en una malla rectangular finita

$$R = \{(i, j) : i = 1, \dots, m, j = 1, \dots, n\} \subset \mathbb{Z}^2$$

Así nuestro proceso está restringido a R y vendrá dado por:

$$X_{ij} = \lambda X_{i-1,j} + \beta X_{i,j-1} - \lambda\beta X_{i-1,j-1} + \eta_{ij}, \quad (i, j) \in R \quad (3.1)$$

Supongamos que las observaciones se obtienen de acuerdo con la siguiente ecuación de observación

$$Y_{ij} = m_{ij}X_{ij} + \epsilon_{ij}, \quad (i, j) \in R \quad (3.2)$$

donde $\epsilon_{ij} \rightsquigarrow N(0, \sigma_\epsilon^2)$ son independientes entre sí y conjuntamente independientes con los η_{ij} , y los coeficientes m_{ij} son constantes conocidas. Asumamos también, por simplicidad, que el proceso X_{ij} tiene media cero (en la práctica, puede interesar la inclusión de una media o función de tendencia determinista).

Introducimos esta ecuación de observación por dos motivos: el primero es que el instrumento de medida puede tener un cierto error no inherente al proceso, y el segundo, que los nodos donde no se disponga de observación pueden ser considerados como con observación igual a cero si $m_{ij} = 0$ y $\epsilon_{ij} = 0$ (este aspecto se desarrollará posteriormente).

La aplicación del algoritmo EM en series temporales (ver Shumway et al. (1982)) requiere el uso de una condición inicial x_0 . En nuestro caso, de acuerdo con la estructura del proceso X , necesitaremos condiciones iniciales asociadas a los bordes Sur y Oeste de la malla. Debido a la propiedad de separabilidad del proceso, parece lógico que estos bordes se pudieran formular de la forma

$$\begin{aligned} X_{00} &\rightsquigarrow N(0, \sigma^2) \\ X_{i0} &= \lambda X_{i-1,0} + \eta_{i0}, & i &= 1, \dots, m \\ X_{0j} &= \beta X_{0,j-1} + \eta_{0j}, & j &= 1, \dots, n \end{aligned}$$

comportándose los ruidos η_{i0} y η_{0j} en principio como los η_{ij} anteriores. Sin embargo, cuando realizamos un estudio asintótico de las predicciones, esta suposición nos conduce a que las varianzas de las predicciones de X_{ij} cuando i y j están suficientemente alejados de 0 tienen un buen comportamiento, pero la varianza de las predicciones de observaciones próximas a los bordes (algún subíndice próximo a 0) es asintóticamente menor de lo que cabría esperar. Para solventar este problema, desarrollamos los puntos de los bordes sobre

todo \mathbb{Z}^2 . Así, desarrollando X_{i0} obtenemos:

$$X_{i0} = \lambda X_{i-1,0} + \beta X_{i,-1} - \lambda\beta X_{i-1,-1} + \eta_{i0}.$$

Si ahora desarrollamos $X_{i,-1}$ obtenemos:

$$X_{i0} = \lambda X_{i-1,0} + \beta(\lambda X_{i-1,-1} + \beta X_{i,-2} - \lambda\beta X_{i-1,-2} + \eta_{i,-1}) - \lambda\beta X_{i-1,-1} + \eta_{i0},$$

con lo que queda:

$$X_{i0} = \lambda X_{i-1,0} + \beta(\beta X_{i,-2} - \lambda\beta X_{i-1,-2} + \eta_{i,-1}) + \eta_{i0}.$$

Si repetimos el proceso desarrollando $X_{i,*}$, observamos que se van cancelando todos los términos $X_{i-1,*}$, salvo $X_{i-1,0}$, y obteniéndose la expresión:

$$X_{i0} = \lambda X_{i-1,0} + \eta_{i0} + \beta\eta_{i,-1} + \beta^2\eta_{i,-2} + \beta^3\eta_{i,-3} + \dots$$

En definitiva, X_{i0} puede ser modelado de la forma:

$$X_{i0} = \lambda X_{i-1,0} + \tilde{\eta}_{i0}, \tag{3.3}$$

donde $\tilde{\eta}_{i0} \rightsquigarrow N(0, \frac{\sigma_{\eta}^2}{1-\beta^2})$, ya que $\tilde{\eta}_{i0}$ se obtiene como combinación lineal de normales independientes.

Razonando de igual forma para X_{0j} obtenemos:

$$X_{0j} = \beta X_{0,j-1} + \tilde{\eta}_{0j}, \tag{3.4}$$

donde $\tilde{\eta}_{0j} \rightsquigarrow N(0, \frac{\sigma_{\eta}^2}{1-\lambda^2})$.

Análogamente puede verse que una exigencia natural es:

$$X_{00} \rightsquigarrow N\left(0, \frac{\sigma_{\eta}^2}{(1-\lambda^2)(1-\beta^2)}\right). \tag{3.5}$$

Además, puede verse que el desarrollo de cada nodo en que estamos interesados es función lineal de ruidos independientes, situados en filas (caso 0j), columnas (caso i0) o cuadrante (caso 00) con intersección vacía. Por esto los ruidos η_{ij} , $\tilde{\eta}_{0j}$ y $\tilde{\eta}_{i0}$ son independientes entre sí y con X_{00} .

Por tanto, nuestro modelo viene dado por las ecuaciones (3.1), (3.2), (3.3), (3.4) y por (3.5), donde todos los ruidos que aparecen son independientes entre sí y con X_{00} .

Podemos llegar a las mismas conclusiones, si definimos nuestro proceso en términos de operadores de retardos. Así, recordemos que los operadores de retardos B_1 y B_2 se definieron como aquellos que verificaban que

$B_1 X_{ij} = X_{i-1,j}$ y $B_2 X_{ij} = X_{i,j-1}$. Por esto el proceso de interés X se puede definir como:

$$(1 - \lambda B_1)(1 - \beta B_2)X_{ij} = \eta_{ij}.$$

El borde horizontal se expresará en función del valor del proceso que está a su izquierda y de una combinación de valores de ruido hacia abajo, resultado de la siguiente expresión:

$$(1 - \lambda B_1)X_{i0} = \frac{1}{1 - \beta B_2} \eta_{i0}.$$

Si desarrollamos el cociente del segundo término obtenemos:

$$\frac{1}{1 - \beta B_2} = 1 + \beta B_2 + \beta^2 B_2^2 + \beta^3 B_2^3 + \dots$$

Por esto se deduce de nuevo que el modelo adopta la expresión:

$$X_{i0} = \lambda X_{i-1,0} + \eta_{i0} + \beta \eta_{i,-1} + \beta^2 \eta_{i,-2} + \beta^3 \eta_{i,-3} + \dots,$$

con lo que X_{i0} puede ser modelado según la ecuación (3.3).

De forma análoga, el borde vertical y la observación situada en el punto (0,0) adoptarán la expresión:

$$(1 - \beta B_2)X_{0j} = \frac{1}{1 - \lambda B_1} \eta_{0j},$$

$$X_{00} = \frac{1}{(1 - \lambda B_1)(1 - \beta B_2)} \eta_{00},$$

obteniéndose los resultados presentados anteriormente.

3.3 Adaptación del algoritmo EM

Nuestro problema de estimación consiste en obtener un valor para los parámetros λ y β , y las varianzas σ_η^2 y σ_ϵ^2 , a partir de los datos observados.

Al ser los ruidos y X_{00} independientes y con distribución normal, el logaritmo de su función de verosimilitud conjunta adopta la expresión:

$$\log L^* \doteq \log f(\eta_{ij}, \epsilon_{i,j}, (i,j) \in R, x_{00}, \tilde{\eta}_{i0}, \tilde{\eta}_{0j}, i = 1, \dots, m, j = 1, \dots, n) =$$

$$\begin{aligned}
 &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n \epsilon_{ij}^2 - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log \sigma_\eta^2 - \\
 &\quad - \frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n \eta_{ij}^2 - \frac{1}{2} \log \frac{\sigma_\eta^2}{(1-\lambda^2)(1-\beta^2)} - \\
 &\quad - \frac{(1-\lambda^2)(1-\beta^2)}{2\sigma_\eta^2} x_{00}^2 - \frac{1}{2} \sum_{i=1}^m \log \frac{\sigma_\eta^2}{1-\beta^2} - \frac{1-\beta^2}{2\sigma_\eta^2} \sum_{i=1}^m \tilde{\eta}_{i0}^2 - \\
 &\quad - \frac{1}{2} \sum_{j=1}^n \log \frac{\sigma_\eta^2}{1-\lambda^2} - \frac{1-\lambda^2}{2\sigma_\eta^2} \sum_{j=1}^n \tilde{\eta}_{0j}^2.
 \end{aligned}$$

Teniendo en cuenta el modelo global (3.1)-(3.2)-(3.3)-(3.4)-(3.5), el logaritmo de la función de verosimilitud de los datos completos viene dado por:

$$\begin{aligned}
 \log L &\doteq \log f(x_{ij}, y_{ij}, (i, j) \in R, x_{00}, x_{i0}, x_{0j}, i = 1, \dots, m, j = 1, \dots, n) = \\
 &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij}x_{ij})^2 - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log \sigma_\eta^2 - \\
 &\quad - \frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \lambda x_{i-1,j} - \beta x_{i,j-1} - \lambda\beta x_{i-1,j-1})^2 - \\
 &\quad - \frac{1}{2} \log \frac{\sigma_\eta^2}{(1-\lambda^2)(1-\beta^2)} - \frac{(1-\lambda^2)(1-\beta^2)}{2\sigma_\eta^2} x_{00}^2 - \\
 &\quad - \frac{1}{2} \sum_{i=1}^m \log \frac{\sigma_\eta^2}{1-\beta^2} - \frac{1-\beta^2}{2\sigma_\eta^2} \sum_{i=1}^m (x_{i0} - \lambda x_{i-1,0})^2 - \\
 &\quad - \frac{1}{2} \sum_{j=1}^n \log \frac{\sigma_\eta^2}{1-\lambda^2} - \frac{1-\lambda^2}{2\sigma_\eta^2} \sum_{j=1}^n (x_{0j} - \beta x_{0,j-1})^2. \tag{3.6}
 \end{aligned}$$

También se puede reformular esta función de la siguiente forma:

$$\begin{aligned}
 \log L &\doteq -\frac{mn}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij}x_{ij})^2 - \\
 &\quad - \frac{mn + m + n + 1}{2} \log \sigma_\eta^2 + \frac{n+1}{2} \log(1-\lambda^2) + \\
 &\quad + \frac{m+1}{2} \log(1-\beta^2) - \frac{1}{2\sigma_\eta^2} (1, \lambda, \lambda^2) \mathbf{A} (1, \beta, \beta^2)', \tag{3.7}
 \end{aligned}$$

donde \mathbf{A} es una matriz 3×3 cuyos términos a_{pq} vienen dados por:

$$\begin{aligned}
 a_{11} &= \sum_{i=0}^m \sum_{j=0}^n x_{ij}^2 \\
 a_{12} &= -2 \sum_{i=1}^m \sum_{j=0}^n x_{ij} x_{i,j-1} \\
 a_{13} &= \sum_{i=0}^m \sum_{j=2}^n x_{i,j-1}^2 \\
 a_{21} &= -2 \sum_{i=1}^m \sum_{j=0}^n x_{ij} x_{i-1,j} \\
 a_{22} &= 2 \sum_{i=1}^m \sum_{j=1}^n (x_{ij} x_{i-1,j-1} + x_{i-1,j} x_{i,j-1}) \\
 a_{23} &= -2 \sum_{i=1}^m \sum_{j=2}^n x_{i,j-1} x_{i-1,j-1} \\
 a_{31} &= \sum_{i=2}^m \sum_{j=0}^n x_{i-1,j}^2 \\
 a_{32} &= -2 \sum_{i=2}^m \sum_{j=1}^n x_{i-1,j} x_{i-1,j-1} \\
 a_{33} &= \sum_{i=2}^m \sum_{j=2}^n x_{i-1,j-1}^2. \tag{3.8}
 \end{aligned}$$

Hay que hacer notar que en el caso de observaciones faltantes la función de log-verosimilitud se ve modificada en sus dos primeros sumandos, ya que el número de ϵ_{ij} a considerar es menor: El primer sumando reflejará el número de observaciones realizadas, en el segundo sumando aparecerán las sumas de los términos $(y_{ij} - m_{ij} x_{ij})^2$ en los nodos donde se dispone de observación, por lo que, a efectos computacionales, se puede considerar la suma en todos los nodos pero introduciendo en los nodos con observaciones faltantes $y_{ij} = 0$ y $m_{ij} = 0$, lo cual proporciona un sumando nulo, que es equivalente a no sumar nada en ese nodo.

El paso E del algoritmo involucra el cálculo de la función:

$$Q(\psi \mid \psi^{(r)}) = E_{\psi^{(r)}}(\log L \mid \mathbf{Y}),$$

donde $\psi = (\lambda, \beta, \sigma_\eta^2, \sigma_\epsilon^2)$, y $\psi^{(r)}$ es el valor estimado de ψ en la iteración r-ésima y donde \mathbf{Y} es el conjunto de observaciones $\{y_{ij} : (i, j) \in R\}$.

Realizando algunos cálculos llegamos a la expresión:

$$\begin{aligned}
 Q(\psi | \psi^{(r)}) \doteq & -\frac{1}{2} \log \frac{\sigma_\eta^2}{(1-\lambda^2)(1-\beta^2)} - \frac{(1-\lambda^2)(1-\beta^2)}{\sigma_\eta^2} [P_{00}^{mn} + x_{00}^{mn}] - \\
 & -\frac{mn}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n [m_{ij}^2 P_{ij}^{mn} + (y_{ij}^{mn} - m_{ij} x_{ij}^{mn})^2] - \\
 & -\frac{mn+m+n}{2} \log \sigma_\eta^2 - \frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n [(x_{ij}^{mn} - \lambda x_{i-1,j}^{mn} - \\
 & -\beta x_{i,j-1}^{mn} + \lambda\beta x_{i-1,j-1}^{mn})^2 + P_{ij}^{mn} + \lambda^2 P_{i-1,j}^{mn} + \beta^2 P_{i,j-1}^{mn} + \\
 & + \lambda^2 \beta^2 P_{i-1,j-1}^{mn} - 2\lambda P_{ij,i-1,j}^{mn} - 2\beta P_{ij,i,j-1}^{mn} + 2\lambda\beta P_{i-1,j,i,j-1}^{mn} + \\
 & + 2\lambda\beta P_{ij,i-1,j-1}^{mn} - 2\lambda^2 \beta P_{i-1,j,i-1,j-1}^{mn} - 2\lambda\beta^2 P_{i,j-1,i-1,j-1}^{mn}] - \\
 & -\frac{1-\beta^2}{2\sigma_\eta^2} \sum_{i=1}^m [P_{i0}^{mn} - 2\lambda P_{i0,i-1,0}^{mn} + \lambda^2 P_{i-1,0}^{mn} + (x_{i0}^{mn} - \lambda x_{i-1,0}^{mn})^2] - \\
 & -\frac{1-\lambda^2}{2\sigma_\eta^2} \sum_{j=1}^n [P_{0j}^{mn} - 2\beta P_{0j,0,j-1}^{mn} + \beta^2 P_{0,j-1}^{mn} + (x_{0j}^{mn} - \beta x_{0,j-1}^{mn})^2] + \\
 & + \frac{m}{2} \log(1-\beta^2) + \frac{n}{2} \log(1-\lambda^2),
 \end{aligned} \tag{3.9}$$

donde $x_{ij}^{mn} = E[x_{ij}/\mathbf{Y}]$, $P_{ij}^{mn} = \text{Var}[x_{ij}/\mathbf{Y}]$, y $P_{ijkl}^{mn} = \text{Cov}[x_{ij}, x_{kl}/\mathbf{Y}]$. Así, podemos observar que la función $Q(\psi | \psi^{(r)})$ se expresa en términos de esperanzas y covarianzas condicionadas.

Esta función se puede reformular usando una expresión matricial como:

$$\begin{aligned}
 Q(\psi | \psi^{(r)}) \doteq & -\frac{mn}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n [m_{ij}^2 P_{ij}^{mn} + (y_{ij}^{mn} - m_{ij} x_{ij}^{mn})^2] - \\
 & -\frac{mn+m+n+1}{2} \log \sigma_\eta^2 + \frac{n+1}{2} \log(1-\lambda^2) + \\
 & + \frac{m+1}{2} \log(1-\beta^2) - \frac{1}{2\sigma_\eta^2} (1, \lambda, \lambda^2) \bar{\mathbf{A}} (1, \beta, \beta^2)' \tag{3.10}
 \end{aligned}$$

donde los coeficientes \bar{a}_{pq} ($p, q = 1, 2, 3$) de la matriz $\bar{\mathbf{A}}$, vienen dados en la expresión (3.13).

El paso M consiste en, dado $\psi^{(r)}$, calcular $\psi^{(r+1)}$ de forma que verifique:

$$Q(\psi^{(r+1)} | \psi^{(r)}) \geq Q(\psi | \psi^{(r)}), \quad \forall \psi.$$

Claramente el vector $\psi^{(r+1)}$ obtenido al maximizar la función Q depende de las esperanzas y varianzas-covarianzas de la variable X condicionadas a \mathbf{Y} . Para el caso de series temporales, Shumway et al. (1982) reúnen en su trabajo una serie de procedimientos (ver Jazwinski (1970) y Stoffer (1982)) basados en el filtrado de Kalman para calcular las cantidades condicionadas que se precisan. En nuestro caso, debido a que el proceso es gaussiano, (\mathbf{X}, \mathbf{Y}) tiene distribución conjunta normal multivariante de dimensión $2nm + m + n + 1$. Así, las esperanzas y varianzas condicionadas vienen dadas por:

$$\Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{Y}, \quad \Sigma_{\mathbf{X}\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1} \Sigma_{\mathbf{Y}\mathbf{X}}, \quad (3.11)$$

donde $\Sigma_{\mathbf{X}\mathbf{X}}$, $\Sigma_{\mathbf{X}\mathbf{Y}}$ y $\Sigma_{\mathbf{Y}\mathbf{Y}}$ representan las matrices de covarianzas de (\mathbf{X}, \mathbf{X}) , (\mathbf{X}, \mathbf{Y}) y (\mathbf{Y}, \mathbf{Y}) , respectivamente.

El cálculo de (3.11) es complicado, ya que implica la inversión de una matriz $mn \times mn$, $\Sigma_{\mathbf{Y}\mathbf{Y}}$. Para subsanar este problema proponemos calcular una base ortogonal del subespacio que generan las observaciones Y_{ij} . A diferencia de como hace Jones (1989), no proyectamos sobre las observaciones que conocemos, lo que nos daría la base

$$\begin{bmatrix} \tilde{y}_1 = y_1 \\ \tilde{y}_2 = y_2 - H_1^2 y_1 \\ \vdots \\ \tilde{y}_T = y_T - H_1^T y_1 - \dots - H_{T-1}^T y_{T-1} \end{bmatrix},$$

sino que aprovechamos las variables que hemos construido en los pasos anteriores para proyectar sobre ellas, obteniendo la base

$$\begin{bmatrix} \tilde{y}_1 = y_1 \\ \tilde{y}_2 = y_2 - H_1^2 \tilde{y}_1 \\ \vdots \\ \tilde{y}_T = y_T - H_1^T \tilde{y}_1 - \dots - H_{T-1}^T \tilde{y}_{T-1} \end{bmatrix}.$$

Por tanto, las matrices que debemos calcular serán:

$$\Sigma_{\mathbf{X}\tilde{\mathbf{Y}}} \Sigma_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} \quad \Sigma_{\mathbf{X}\mathbf{X}} - \Sigma_{\mathbf{X}\tilde{\mathbf{Y}}} \Sigma_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^{-1} \Sigma_{\tilde{\mathbf{Y}}\mathbf{X}}.$$

La ventaja de nuestro método es que además de obtener una base de variables incorreladas, el cálculo de las varianzas de las mismas es más simple, ya que $\tilde{y}_i = y_i - H_1^i \tilde{y}_1 - \dots - H_{i-1}^i \tilde{y}_{i-1}$ y \tilde{y}_j son incorreladas, para $j < i$, por lo que no aparecen las covarianzas cruzadas entre dichas variables, que sí aparecen en el procedimiento de Jones (1989).

Los coeficientes H_r^s están, por tanto, definidos de la forma

$$H_r^s = \frac{C_{rs}^{\tilde{Y}Y}}{C_{rr}^{\tilde{Y}Y}},$$

donde

$$C_{rs}^{\tilde{Y}Y} = \text{Cov}[\tilde{y}_r, y_s], \quad r, s = 1, \dots, T.$$

En principio cabría pensar que en el denominador de H_r^s debería de aparecer la varianza de \tilde{y}_r , ya que en realidad lo que se hace para calcular \tilde{y}_s es una regresión de y_s sobre las variables ortogonales que se han construido anteriormente ($1 \leq r < s$) y posteriormente restar a y_s la regresión obtenida. En esta regresión, como las variables tienen esperanza cero, los coeficientes de regresión tienen la expresión $\text{Cov}(\tilde{y}_r, y_s) / \text{Var}(\tilde{y}_r)$. Se puede observar que esta cantidad es igual a $C_{rr}^{\tilde{Y}Y}$, ya que $\text{Var}(\tilde{y}_r) = \text{Cov}(\tilde{y}_r, \tilde{y}_r) = \text{Cov}(\tilde{y}_r, y_r - H_1^r \tilde{y}_1 - \dots - H_{r-1}^r \tilde{y}_{r-1})$ y, puesto que \tilde{y}_r es incorrelado con \tilde{y}_i para $i < r$, la covarianza anterior resulta ser igual a $\text{Cov}(\tilde{y}_r, y_r)$.

Las covarianzas necesarias para el cálculo de H_r^s se pueden calcular de la siguiente forma recursiva:

$$C_{1s}^{\tilde{Y}Y} = C_{1s}^{YY},$$

$$C_{rs}^{\tilde{Y}Y} = C_{rs}^{YY} - \sum_{t=1}^{r-1} \frac{C_{tr}^{\tilde{Y}Y} C_{ts}^{\tilde{Y}Y}}{C_{rr}^{\tilde{Y}Y}}, \quad \text{para } r \geq 2,$$

donde $C_{rs}^{YY} = \text{Cov}(y_r, y_s)$ es igual a

$$C_{rs}^{YY} = m_r m_s C_{rs}^{XX} + \delta_{rs} \sigma_\epsilon^2, \quad r, s = 1, \dots, T.$$

Como se vio anteriormente

$$C_{rs}^{XX} = C_{(i,j),(k,l)}^{XX} = \frac{\lambda^{|i-k|} \beta^{|j-l|} \sigma_\eta^2}{(1-\lambda^2)(1-\beta^2)},$$

donde r se corresponde con (i, j) , y s con (k, l) .

Análogamente, la matriz $\Sigma_{\mathbf{X}\tilde{\mathbf{Y}}}$ se puede calcular obteniendo las covarianzas $C_{rs}^{X\tilde{Y}} = \text{Cov}[x_r, \tilde{y}_s]$ de forma recursiva a partir de las expresiones

$$C_{r1}^{X\tilde{Y}} = m_1 C_{r1}^{XX}$$

$$C_{rs}^{X\tilde{Y}} = m_s C_{rs}^{XX} - \sum_{t=1}^{s-1} H_t^s C_{rt}^{X\tilde{Y}}, \quad \text{para } s \geq 2.$$

3.4 Estimación y alisamiento

Como se comentó anteriormente, el valor de $\psi^{(r+1)}$ se calcula iterativamente, de forma que

$$Q(\psi^{(r+1)} | \psi^{(r)}) \geq Q(\psi | \psi^{(r)}), \quad \forall \psi.$$

Es decir, debemos encontrar el valor $\psi^{(r+1)}$ tal que

$$Q(\psi^{(r+1)} | \psi^{(r)}) = \max_{\psi} Q(\psi | \psi^{(r)}),$$

el cual se obtendrá maximizando la función Q , cuya expresión viene dada por la ecuación (3.9). Para maximizar la función anterior, la derivamos con respecto a los parámetros λ , β , σ_η^2 y σ_ϵ^2 . Igualando las derivadas a cero obtenemos cuatro ecuaciones que proporcionan los valores de λ , β , σ_η^2 y σ_ϵ^2 . Estas ecuaciones son:

$$\begin{aligned} & \bar{a}_{11} + \bar{a}_{31}\lambda^2 + \bar{a}_{13}\beta^2 + \bar{a}_{33}\lambda^2\beta^2 + \bar{a}_{21}\lambda + \bar{a}_{12}\beta + \bar{a}_{22}\lambda\beta + \bar{a}_{32}\lambda^2\beta + \bar{a}_{23}\lambda\beta^2 - \\ & \quad - (m+1)(n+1)\sigma_\eta^2 = 0 \\ & (1-\lambda^2)[\bar{a}_{21} + 2\bar{a}_{31}\lambda + \bar{a}_{22}\beta + 2\bar{a}_{33}\lambda\beta^2 + 2\bar{a}_{32}\lambda\beta + \bar{a}_{23}\beta^2] + 2\lambda(n+1)\sigma_\eta^2 = 0 \\ & (1-\beta^2)[\bar{a}_{12} + 2\bar{a}_{13}\beta + \bar{a}_{22}\lambda + 2\bar{a}_{33}\lambda^2\beta + 2\bar{a}_{23}\lambda\beta + \bar{a}_{32}\lambda^2] + 2\beta(m+1)\sigma_\eta^2 = 0. \\ & \sigma_\epsilon^2 = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [m_{ij} P_{ij}^{mn} + (y_{ij} - m_{ij} x_{ij}^{mn})]. \end{aligned} \tag{3.12}$$

Los coeficientes \bar{a}_{pq} vienen dados por:

$$\begin{aligned}
 \bar{a}_{11} &= \sum_{i=0}^m \sum_{j=0}^n x_{ij}^{mn2} + \sum_{i=0}^m \sum_{j=0}^n P_{ij}^{mn} \\
 \bar{a}_{12} &= -2 \sum_{i=0}^m \sum_{j=1}^n x_{ij}^{mn} x_{i,j-1}^{mn} - 2 \sum_{i=0}^m \sum_{j=1}^n P_{ij,i,j-1}^{mn} \\
 \bar{a}_{13} &= \sum_{i=0}^m \sum_{j=2}^n x_{i,j-1}^{mn2} + \sum_{i=0}^m \sum_{j=2}^n P_{i,j-1}^{mn} \\
 \bar{a}_{21} &= -2 \sum_{i=1}^m \sum_{j=0}^n x_{ij}^{mn} x_{i-1,j}^{mn} - 2 \sum_{i=1}^m \sum_{j=0}^n P_{ij,i-1,j}^{mn} \\
 \bar{a}_{22} &= 2 \sum_{i=1}^m \sum_{j=1}^n (x_{ij}^{mn} x_{i-1,j-1}^{mn} + x_{i-1,j}^{mn} x_{i,j-1}^{mn}) + 2 \sum_{i=1}^m \sum_{j=1}^n (P_{ij,i-1,j-1}^{mn} + P_{i-1,j,i,j-1}^{mn}) \\
 \bar{a}_{31} &= \sum_{i=2}^m \sum_{j=0}^n x_{i-1,j}^{mn2} + \sum_{i=2}^m \sum_{j=0}^n P_{i-1,j}^{mn} \\
 \bar{a}_{32} &= -2 \sum_{i=2}^m \sum_{j=1}^n x_{i-1,j}^{mn} x_{i-1,j-1}^{mn} - 2 \sum_{i=2}^m \sum_{j=1}^n P_{i-1,j,i-1,j-1}^{mn} \\
 \bar{a}_{23} &= -2 \sum_{i=1}^m \sum_{j=2}^n x_{i,j-1}^{mn} x_{i-1,j-1}^{mn} - 2 \sum_{i=1}^m \sum_{j=2}^n P_{i,j-1,i-1,j-1}^{mn} \\
 \bar{a}_{33} &= \sum_{i=2}^m \sum_{j=2}^n x_{i-1,j-1}^{mn2} + \sum_{i=2}^m \sum_{j=2}^n P_{i-1,j-1}^{mn}
 \end{aligned} \tag{3.13}$$

Eliminando las constantes que no dependen de ψ , el valor $Q(\psi^{(r+1)} | \psi^{(r)})$ adopta la expresión

$$\begin{aligned}
 Q(\psi^{(r+1)} | \psi^{(r)}) \approx & -\frac{mn + m + n + 1}{2} \log \sigma_n^{2(r+1)} + \frac{n+1}{2} \log(1 - \lambda^{2(r+1)}) + \\
 & + \frac{m+1}{2} \log(1 - \beta^{2(r+1)}) - \frac{mn}{2} \log \sigma_\epsilon^{2(r+1)}
 \end{aligned}$$

(el factor mn en el último término se sustituirá, en el caso de datos omitidos, por el número de nodos observados).

En resumen, el procedimiento consiste en lo siguiente: En primer lugar, partimos de un valor inicial de los parámetros $\psi(0)$ y, a partir de este, calculamos las esperanzas y covarianzas de \mathbf{X} , y mediante éstas las de \mathbf{Y} . Posteriormente, se construye la "base ortogonal" $\tilde{\mathbf{Y}}$ y con ésta, las esperanzas y

covarianzas condicionadas x_{ij}^{mn} y P_{ijkl}^{mn} . Una vez obtenidas estas cantidades, calculamos la función $Q(\psi | \psi(0))$. Maximizamos esta función y obtenemos $\psi(1)$, con lo que se acaba la primera iteración. Posteriormente, se repite el proceso sustituyendo $\psi(0)$ por $\psi(1)$, y así sucesivamente.

Por último, quedaría por ver cómo se elige la condición inicial $\psi(0)$, ya que de ésta depende, en cierta medida, el número de iteraciones hasta que el algoritmo converja, obviamente dependiendo también de la forma de la función de verosimilitud. Es claro que si la condición inicial está muy lejos del verdadero valor del parámetro, el número de iteraciones será mayor que si la condición inicial está cercana. Para proponer la condición inicial, vamos a considerar las observaciones ordenadas mediante el orden lexicográfico, es decir, $y_{11}, y_{21}, \dots, y_{n1}, y_{12}, y_{22}, \dots, y_{n2}, \dots, y_{1n}, \dots, y_{nn}$. Si consideramos estas observaciones como una serie temporal, podemos observar que el comportamiento de esta serie, si suponemos de partida un modelo $AR(1) \times AR(1)$, es autorregresivo de orden 1 entre $y_{1j}, y_{2j}, \dots, y_{mj}$ y además existe también dependencia autorregresiva de primer orden entre $y_{ij}, y_{i+1,j}, y_{i+2,j}, \dots$, por lo que la serie completa se podrá modelizar adecuadamente mediante un proceso con estructura autorregresiva de orden 1 ordinaria y estructura autorregresiva de orden 1 estacional con estacionariedad de orden n . Se pueden proponer también valores de σ_η^2 y σ_ϵ^2 a partir de la varianza estimada obtenida al ajustar el modelo anterior. (Hay que hacer notar que el procedimiento anterior conlleva la introducción de un error de ajuste, debido al efecto de la yustaposición unidimensional de nodos del borde Este con nodos del borde Oeste).

3.4.1 Problema de identificación

Si observamos el modelo dado por (3.1) y (3.2) y tenemos en cuenta que en la mayoría de los casos las constantes m_{ij} sólo tomarán los valores 0 (si es observación omitida) o 1 (si no lo es), nos damos cuenta de que eventualmente puede ser difícil distinguir entre los ruidos η y ϵ . De acuerdo con dichas ecuaciones, las observaciones responden al siguiente esquema:

$$\begin{aligned} Y_{ij} &= \lambda X_{i-1,j} + \beta X_{i,j-1} - \lambda\beta X_{i-1,j-1} + \eta_{ij} + \epsilon_{ij} & \text{si } m_{ij} = 1 \\ Y_{ij} &= 0 & \text{si } m_{ij} = 0 \end{aligned}$$

En esta situación algunos autores (p.e. ver Katayama (1979)) fijan el valor de la varianza σ_ϵ^2 . Sin embargo, esta solución no parece muy satisfactoria como forma general de proceder. Si transformamos nuestro proceso $AR(1) \times AR(1)$

en un proceso $MA(\infty) \times MA(\infty)$ desarrollándolo sobre \mathbb{Z}^2 , éste tendrá la siguiente forma:

$$X_{ij} = \frac{1}{1 - \lambda B_1} \frac{1}{1 - \beta B_2} \eta_{ij} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \lambda^k \beta^l \eta_{i-k, j-l}.$$

La ecuación de observación adopta ahora la expresión:

$$Y_{ij} = m_{ij} \left(\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \lambda^k \beta^l \eta_{i-k, j-l} \right) + \epsilon_{ij}.$$

Se puede observar que si los m_{ij} son diferentes entre sí, cabe esperar que podamos discriminar la variación (σ_ϵ^2 y σ_η^2) procedente de ϵ y η , ya que en cada nodo tendremos una ponderación distinta de los valores del ruido η . También se observa que la influencia que tienen los inputs de ruido η sobre Y tiene un decaimiento exponencial por lo que si alguno de los parámetros es “bajo”, o los dos, la doble sumatoria anterior tendrá pocos términos “significativos” y esto hará que no podamos discriminar bien entre σ_ϵ^2 y σ_η^2 . Por tanto, las situaciones que, en general, se pueden presentar son las siguientes:

1. Si las constantes m_{ij} son distintas entre sí se podrán estimar bien los cuatro parámetros conjuntamente.
2. Si ambos parámetros, λ y β , son “grandes”, también se estimarán conjuntamente los cuatro parámetros.
3. Si no ocurre ninguna de las dos situaciones anteriores, se intentará fijar el valor de σ_ϵ^2 , con lo que ahora sólo hay que estimar $\psi^* = (\lambda, \beta, \sigma_\eta^2)$. Si esto no fuera posible, se prescindirá del ruido ϵ en la ecuación de observación (lo que equivale a prescindir de dicha ecuación desde el principio).

Teniendo en cuenta estas consideraciones proponemos finalmente el siguiente procedimiento: En primer lugar, estimar los parámetros $\psi = (\lambda, \beta, \sigma_\eta^2, \sigma_\epsilon^2)$. Si se observa que la varianza σ_ϵ^2 se hace muy pequeña, proceder como en 3, estimando sólo tres parámetros $\psi^* = (\lambda, \beta, \sigma_\eta^2)$.

3.4.2 Error de predicción

Una vez que hemos estimado el valor en los nodos, alisado si se disponía de observación en éste o estimado si no se disponía de la misma, podemos calcular el error cuadrático medio para la predicción.

Hemos estimado X_{ij} mediante su esperanza condicionada a las observaciones de que se dispone \mathbf{Y} que notamos x_{ij}^{mn} . Ésta adopta la expresión

$$x_{ij}^{mn} = \sum_{t=1}^T \frac{\text{Cov}(X_{ij}, \tilde{Y}_t)}{\text{Var}\tilde{Y}_t} \tilde{y}_t,$$

por lo que el error cuadrático medio adopta la expresión

$$E[(x_{ij} - x_{ij}^{mn})^2 | \mathbf{Y}] = \sigma_X^2 - \sum_{t=1}^T \frac{(\text{Cov}(X_{ij}, \tilde{Y}_t))^2}{\text{Var}\tilde{Y}_t}.$$

3.5 El algoritmo ECM

Estudiemos con detenimiento los cálculos necesarios para la aplicación del algoritmo EM en cada iteración. En primer lugar partimos de unos valores iniciales de los parámetros, que generalmente variarán de iteración en iteración, sobre los cuales construimos nuestra base ortogonal. Para ello debemos de calcular todas las covarianzas cruzadas entre los distintos retardos entre x_{ij} y x_{kl} , y a partir de estas, de forma recursiva, la base ortogonal, las varianzas de dicha base y las covarianzas entre esta base y los x_{ij} . Una vez que tenemos calculadas todas estas covarianzas, calculamos las esperanzas y covarianzas condicionadas que aparecen en la función Q . Posteriormente calculamos las sumas $\bar{a}_{11}, \bar{a}_{12}, \bar{a}_{13}, \bar{a}_{21}, \bar{a}_{22}, \bar{a}_{23}, \bar{a}_{31}, \bar{a}_{32}, \bar{a}_{33}$. Ahora "simplemente" nos queda resolver el sistema no lineal (3.12) o, en su caso, maximizar directamente la función Q .

En ambos casos el problema no es fácil de resolver computacionalmente. Para maximizar la función hay que emplear métodos numéricos. Se intentó utilizar alguna de las rutinas de la librería IMSL para este fin, pero al ser la función Q demasiado plana en entornos del máximo los algoritmos ensayados de esta librería no convergían. Una primera alternativa consistió en buscar el máximo en una malla suficientemente fina para los parámetros λ y β , ya que de la primera ecuación de (3.12) podemos calcular directamente el valor de σ_η^2 , y de la cuarta el de σ_ϵ^2 , que junto a λ y β hacen máxima la función Q . Sin embargo, este método conlleva un gran número de evaluaciones de la función Q y, por tanto, un número de operaciones muy elevado. Observemos que si buscamos λ y β entre 0 y 1 con dos cifras decimales de precisión, necesitamos 101^2 evaluaciones, las cuales aumentan exponencialmente a medida que se eleva

la precisión. No obstante, por la simplicidad de dicho método, se implementó obteniéndose resultados muy satisfactorios.

Otra posibilidad sería implementar nosotros directamente un algoritmo para el cálculo del máximo de una función de varias variables; por ejemplo, el del gradiente conjugado. Sin embargo, debido a las buenas aproximaciones que obteníamos en la búsqueda en malla, y a que no consideramos que éste sea (con dos cifras decimales) demasiado lento, desechamos esta idea dirigiéndonos a una tercera posibilidad: **el algoritmo ECM**.

Como hemos presentado con anterioridad, en el algoritmo ECM las maximizaciones se hacen condicionadas a ciertas funciones g_s en cada paso CM. En nuestro caso, el conjunto de funciones, G , que nos dan las restricciones de búsqueda ha sido definido de la siguiente forma:

$$\begin{aligned} g_1(\psi) &= (\beta, \sigma_\eta^2, \sigma_\epsilon^2) \\ g_2(\psi) &= (\lambda, \sigma_\eta^2, \sigma_\epsilon^2) \\ g_3(\psi) &= (\lambda, \beta, \sigma_\epsilon^2) \\ g_4(\psi) &= (\lambda, \beta, \sigma_\eta^2) \quad . \end{aligned} \tag{3.14}$$

En el caso en que sólo estimamos tres parámetros $\psi^* = (\lambda, \beta, \sigma_\eta^2)$, las funciones g_s vendrán dadas por:

$$\begin{aligned} g_1(\psi^*) &= (\beta, \sigma_\eta^2) \\ g_2(\psi^*) &= (\lambda, \sigma_\eta^2) \\ g_3(\psi^*) &= (\lambda, \beta) \quad . \end{aligned} \tag{3.15}$$

Es decir, en el primer paso maximizamos sólo en λ , en el segundo sólo en β , en el tercero en σ_η^2 y en el cuarto, si éste se realiza, sólo en σ_ϵ^2 . Claramente estas tres funciones hacen que G sea un conjunto rellenante del espacio en cualquier punto ψ (o ψ^*). Hay que hacer notar además que ahora las maximizaciones son de funciones uniparamétricas. En el primer paso CM β , σ_η^2 y σ_ϵ^2 están fijos y los siguientes pasos CM son análogos cambiando los parámetros que permanecen inmóviles. Para los dos primeros pasos CM utilizamos la rutina UVMID de la librería IMSL, para la cual se proporciona la función Q y su derivada. Calculamos el tercer y cuarto paso CM directamente, ya que se puede despejar este máximo de la primera y última ecuaciones del sistema (3.12).

3.6 Modelo AR(2) \times AR(1)

Suponemos un modelo separable en el sentido de Martin (1979), pero ahora la covarianza en sentido horizontal viene dada por un proceso AR(2), W_t , y la vertical por un proceso AR(1), W'_t . Las funciones de autocovarianzas de dichos procesos (ver Priestley (1981) Vol.1), cuando las raíces del polinomio de retardos tienen módulo mayor que 1, son:

$$(1 - \lambda_1 B)(1 - \lambda_2 B)W_t = \zeta_t,$$

$$\text{Cov}(W_t, W_k) = \frac{(1 - \lambda_2^2)\lambda_1^{|t-k|+1} - (1 - \lambda_1^2)\lambda_2^{|t-k|+1}}{(1 - \lambda_1\lambda_2)(\lambda_1 - \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}\sigma_\zeta^2.$$

$$(1 - \beta B)W'_t = \zeta_t,$$

$$\text{Cov}(W'_t, W'_k) = \frac{\beta^{|t-k|}\sigma_\zeta^2}{1 - \beta^2}.$$

Combinando los dos polinomios de retardos, uno en el índice horizontal (B_1) y otro en el vertical (B_2), obtenemos

$$(1 - (\lambda_1 + \lambda_2)B_1 + \lambda_1\lambda_2 B_1^2)(1 - \beta B_2) = 1 - (\lambda_1 + \lambda_2)B_1 + \lambda_1\lambda_2 B_1^2 - \beta B_2 + \beta(\lambda_1 + \lambda_2)B_1 B_2 - \lambda_1\lambda_2\beta B_1^2 B_2,$$

por lo que el modelo para X adopta la expresión

$$X_{ij} = (\lambda_1 + \lambda_2)X_{i-1,j} - \lambda_1\lambda_2 X_{i-2,j} + \beta X_{i,j-1} - \beta(\lambda_1 + \lambda_2)X_{i-1,j-1} + \lambda_1\lambda_2\beta X_{i-2,j-1} + \eta_{ij}, \quad (3.16)$$

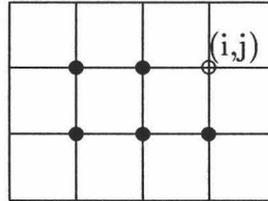
donde $\eta_{ij} \rightsquigarrow N(0, \sigma_\eta^2)$ son independientes y los coeficientes λ_1 , λ_2 y β son constantes. Por tanto, la covarianza entre las variables X_{ij} y X_{kl} viene dada por

$$C_{(i,j),(k,l)}^{XX} = \frac{\beta^{|j-l|} [(1 - \lambda_2^2)\lambda_1^{|i-k|+1} - (1 - \lambda_1^2)\lambda_2^{|i-k|+1}]}{(1 - \beta^2)(1 - \lambda_1\lambda_2)(\lambda_1 - \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}\sigma_\eta^2, \quad (3.17)$$

siendo pues la varianza del proceso

$$\text{Var}X = \frac{(1 + \lambda_1\lambda_2)}{(1 - \beta^2)(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}\sigma_\eta^2.$$

Ahora cada nodo está en relación con otros cinco nodos, que se encuentran situados según se muestra el siguiente dibujo.



De forma análoga a como dedujimos la estructura de las condiciones iniciales para el modelo $AR(1) \times AR(1)$ restringido a una malla finita R , lo hacemos para el que nos ocupa ahora. Para ello desarrollamos los bordes sobre \mathbb{Z}^2 . En primer lugar nos centramos en el borde horizontal, es decir, desarrollamos X_{i0} :

$$X_{i0} = (\lambda_1 + \lambda_2)X_{i-1,0} - \lambda_1\lambda_2X_{i-2,0} + \beta X_{i,-1} - \beta(\lambda_1 + \lambda_2)X_{i-1,-1} + \lambda_1\lambda_2\beta X_{i-2,-1} + \eta_{i0};$$

ahora desarrollamos $X_{i,-1}$, obteniendo

$$X_{i0} = (\lambda_1 + \lambda_2)X_{i-1,0} - \lambda_1\lambda_2X_{i-2,0} + \beta[(\lambda_1 + \lambda_2)X_{i-1,-1} - \lambda_1\lambda_2X_{i-2,-1} + \beta X_{i,-2} - \beta(\lambda_1 + \lambda_2)X_{i-1,-2} + \lambda_1\lambda_2\beta X_{i-2,-2} + \eta_{i,-1}] - \beta(\lambda_1 + \lambda_2)X_{i-1,-1} + \lambda_1\lambda_2\beta X_{i-2,-1} + \eta_{i0}.$$

Simplificamos los sumandos que aparecen con signo opuesto y obtenemos

$$X_{i0} = (\lambda_1 + \lambda_2)X_{i-1,0} - \lambda_1\lambda_2X_{i-2,0} + \beta^2 X_{i,-2} - \beta^2(\lambda_1 + \lambda_2)X_{i-1,-2} + \lambda_1\lambda_2\beta^2 X_{i-2,-2} + \beta\eta_{i,-1} + \eta_{i0}.$$

Análogamente desarrollamos $X_{i,-2}$ obteniendo finalmente la expresión

$$X_{i0} = (\lambda_1 + \lambda_2)X_{i-1,0} - \lambda_1\lambda_2X_{i-2,0} + \beta^3 X_{i,-3} - \beta^3(\lambda_1 + \lambda_2)X_{i-1,-3} + \lambda_1\lambda_2\beta^3 X_{i-2,-3} + \beta^2\eta_{i,-2} + \beta\eta_{i,-1} + \eta_{i0}.$$

Así, en general, obtenemos:

$$\begin{aligned}
(1 - \lambda_1 B_1)(1 - \lambda_2 B_1)X_{i0} &= \beta^n X_{i,-n} - \beta^n(\lambda_1 + \lambda_2)X_{i-1,-n} + \\
&+ \lambda_1 \lambda_2 \beta^n X_{i-2,-n} + \beta^{n-1} \eta_{i,-n+1} + \\
&+ \beta^{n-2} \eta_{i,-n+2} + \cdots + \eta_{i0}.
\end{aligned}$$

Debido a que $|\beta| < 1$, $\beta^n \rightarrow 0$ cuando n tiende a infinito. Por tanto, en el borde horizontal inferior de R las condiciones iniciales se puede modelizar de acuerdo con una estructura autorregresiva de orden 2, dada por

$$(1 - \lambda_1 B_1)(1 - \lambda_2 B_1)X_{i0} = \tilde{\eta}_{i0}, \quad (3.18)$$

donde λ_1 y λ_2 son los coeficientes que aparecen en el modelo general (3.16) y $\tilde{\eta}_{i0} \rightsquigarrow N(0, \frac{\sigma_\eta^2}{1-\beta^2})$ independientemente de los η_{ij} asociados a la malla, ya que es combinación lineal de valores de ruido independientes con éstos últimos.

Hagamos un razonamiento análogo, ahora considerando las dos columnas adyacentes al borde vertical izquierdo, comenzando por $i = -1$.

$$\begin{aligned}
X_{-1j} &= (\lambda_1 + \lambda_2)X_{-2,j} - \lambda_1 \lambda_2 X_{-3,j} + \beta X_{-1,j-1} - \\
&- \beta(\lambda_1 + \lambda_2)X_{-2,j-1} + \lambda_1 \lambda_2 \beta X_{-3,j-1} + \eta_{-1j}.
\end{aligned}$$

Ahora desarrollamos la variable $X_{-2,j}$:

$$\begin{aligned}
(1 - \beta B_2)X_{-1j} &= (\lambda_1 + \lambda_2)[(\lambda_1 + \lambda_2)X_{-3,j} - \lambda_1 \lambda_2 X_{-4,j} + \\
&+ \beta X_{-2,j-1} - \beta(\lambda_1 + \lambda_2)X_{-3,j-1} + \lambda_1 \lambda_2 \beta X_{-4,j-1} + \\
&+ \eta_{-2j}] - \lambda_1 \lambda_2 X_{-3,j} - \beta(\lambda_1 + \lambda_2)X_{-2,j-1} + \\
&+ \lambda_1 \lambda_2 X_{-3,j-1} + (\lambda_1 + \lambda_2)\eta_{-2,j} + \eta_{-1,j} = \\
&= [(\lambda_1 + \lambda_2)^2 - \lambda_1 \lambda_2]X_{-3,j} - \\
&- \beta[(\lambda_1 + \lambda_2)^2 - \lambda_1 \lambda_2]X_{-3,j-1} - \\
&- (\lambda_1 + \lambda_2)\lambda_1 \lambda_2 X_{-4,j} + \beta(\lambda_1 + \lambda_2)\lambda_1 \lambda_2 X_{-4,j-1} + \\
&+ (\lambda_1 + \lambda_2)\eta_{-2,j} + \eta_{-1,j}.
\end{aligned}$$

Análogamente desarrollando la variable $X_{-3,j}$ se obtiene

$$\begin{aligned}
 (1 - \beta B_2)X_{-1j} = & [(\lambda_1 + \lambda_2)^3 - 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)]X_{-4,j} - \\
 & -\beta[(\lambda_1 + \lambda_2)^3 - 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)]X_{-4,j-1} - \\
 & -\lambda_1\lambda_2[(\lambda_1 + \lambda_2)^2 - \lambda_1\lambda_2]X_{-5,j} + \\
 & +\beta\lambda_1\lambda_2[(\lambda_1 + \lambda_2)^2 - \lambda_1\lambda_2]X_{-5,j-1} + \\
 & +[(\lambda_1 + \lambda_2)^2 - \lambda_1\lambda_2]\eta_{-3,j} + (\lambda_1 + \lambda_2)\eta_{-2,j} + \eta_{-1,j}.
 \end{aligned}$$

Se puede observar que los coeficientes asociados a cada valor del ruido tienen la siguiente estructura.

término en l	coeficiente
-1	1
-2	$\lambda_1 + \lambda_2$
-3	$(\lambda_1 + \lambda_2)^2 - \lambda_1\lambda_2$
-4	$(\lambda_1 + \lambda_2)^3 - 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)$
-5	$(\lambda_1 + \lambda_2)^4 - 3\lambda_1\lambda_2(\lambda_1 + \lambda_2)^2 + \lambda_1^2\lambda_2^2$
⋮	⋮

Es decir, dados los coeficientes de los términos -1 y -2, los demás se pueden calcular de manera recursiva, de acuerdo con la siguiente expresión (notamos por $a(l)$ el coeficiente del valor del ruido $\eta_{-i,j}$):

$$a(l + 2) = (\lambda_1 + \lambda_2)a(l + 1) - \lambda_1\lambda_2a(l). \quad (3.19)$$

Con esta nueva notación se observa que

$$\begin{aligned}
 (1 - \beta B_2)X_{-1j} = & a(n)X_{-n,j} - \beta a(n)X_{-n,j-1} - \lambda_1\lambda_2a(n-1)X_{-n,j} + \\
 & +\beta\lambda_1\lambda_2a(n-1)X_{-i-2,j-1} + \sum_{l=1}^{n-1} a(l)\eta_{-l,j} = \\
 = & a(n)[X_{-n,j} - \beta X_{-n,j-1}] - \\
 & -\lambda_1\lambda_2a(n-1)[X_{-n,j} - \beta X_{-i-2,j-1}] + \sum_{l=1}^{n-1} a(l)\eta_{-l,j}.
 \end{aligned}$$

Si consideramos una serie temporal W_t con estructura AR(2) y hacemos un desarrollo análogo al anterior, con la definición de $a(l)$ obtenemos:

$$W_t = a(n)W_{t-n+1} - \lambda_1\lambda_2a(n-1)W_{t-n} + \sum_{l=1}^{n-1} a(l)\zeta_{t-l+1}.$$

Finalmente, la expansión MA(∞) de W_t es:

$$W_t = \sum_{l=1}^{\infty} a(l)\zeta_{t+1-l},$$

Por tanto, en general, el proceso se puede representar de la forma

$$X_{ij} = \beta X_{i,j-1} + \sum_{l=1}^{\infty} a(l)\eta_{i-l+1,j},$$

que en el caso $i = 0$ e $i = -1$ adopta la expresión

$$X_{0j} = \beta X_{0,j-1} + \sum_{l=1}^{\infty} a(l)\eta_{1-l,j}$$

$$X_{-1,j} = \beta X_{-1,j-1} + \sum_{l=1}^{\infty} a(l)\eta_{-l,j}.$$

Definiendo

$$\tilde{\eta}_{0,j} = \sum_{l=1}^{\infty} a(l)\eta_{1-l,j}$$

$$\tilde{\eta}_{-1,j} = \sum_{l=1}^{\infty} a(l)\eta_{-l,j},$$

las condiciones iniciales asociadas al borde vertical Oeste en $i = -1$ e $i = 0$ se pueden modelizar de acuerdo con la siguientes expresiones:

$$(1 - \beta B_2)X_{-1,j} = \tilde{\eta}_{-1,j}, \quad (3.20)$$

$$(1 - \beta B_2)X_{0,j} = \tilde{\eta}_{0,j}. \quad (3.21)$$

Estos nuevos ruidos son independientes de los anteriores (η_{ij} de la malla y $\tilde{\eta}_{i0}$) ya que se obtienen como combinaciones lineales de valores de ruido independientes con los anteriores. También son independientes entre sí, salvo cuando los dos poseen el mismo índice j . En este caso su covarianza viene dada por

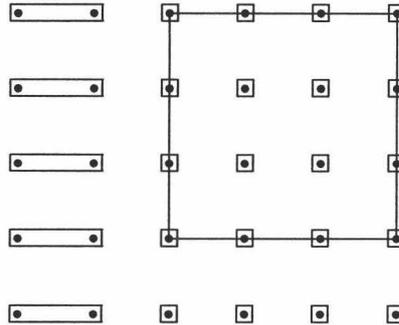
$$\text{Cov}(\tilde{\eta}_{-1,j}, \tilde{\eta}_{0,j}) = \text{Cov}\left(\sum_{i=1}^{\infty} a(i)\eta_{1-i,j}, \sum_{i=1}^{\infty} a(i)\eta_{-i,j}\right) = \frac{(\lambda_1 + \lambda_2)\sigma_{\eta}^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)},$$

ya que se trata de la covarianza entre dos variables de una misma serie con estructura AR(2) con retardo 1.

Así como en el caso $AR(1) \times AR(1)$ necesitábamos una condición inicial X_{00} además de los dos bordes de condiciones iniciales, en este caso necesitamos considerar la distribución de la condición inicial bidimensional $(X_{00}, X_{-1,0})'$. Esta distribución es, claramente, normal bivalente con medias marginales nulas y matriz de covarianzas dada por la estructura de la malla de partida, es decir:

$$\sigma_x^2 \begin{pmatrix} 1 & \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 \lambda_2} \\ \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 \lambda_2} & 1 \end{pmatrix} = \begin{pmatrix} \frac{(1 + \lambda_1 \lambda_2) \sigma_\eta^2}{(1 - \beta^2)(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} & \frac{(\lambda_1 + \lambda_2) \sigma_\eta^2}{(1 - \beta^2)(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \\ \frac{(\lambda_1 + \lambda_2) \sigma_\eta^2}{(1 - \beta^2)(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} & \frac{(1 + \lambda_1 \lambda_2) \sigma_\eta^2}{(1 - \beta^2)(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \end{pmatrix}.$$

Así, nuestro modelo general puede formularse de acuerdo con las ecuaciones (3.16)-(3.18)-(3.20)-(3.21) con todos los ruidos independientes entre sí salvo $\tilde{\eta}_{-1}$, con $\tilde{\eta}_0$, además de suponer que $(X_{00}, X_{-1,0})'$ tiene distribución normal bivalente, con vector de medias nulo y matriz de covarianzas dada anteriormente, además de ser independiente con todos los ruidos anteriores. Por tanto, para calcular la verosimilitud hacemos uso de la distribución normal de los ruidos y condiciones iniciales, y de la estructura de independencia que se puede observar en el siguiente dibujo.



donde los cuadrados hacen referencia a 'inputs' independientes. Los del cuadrado 4×4 hacen referencia a η_{ij} (malla); los de abajo hacen referencia a $\tilde{\eta}_{i0}$; Los rectángulos de la izquierda, salvo el inferior, a $(\tilde{\eta}_{0j}, \tilde{\eta}_{-1,j})'$, y el rectángulo de la esquina inferior izquierda, a $(X_{00}, X_{-1,0})'$. Las distribuciones de estas variables son:

$$\eta_{ij} \rightsquigarrow N(0, \sigma_\eta^2)$$

$$\begin{aligned}
\tilde{\eta}_{i0} &\rightsquigarrow N\left(0, \frac{\sigma_\eta^2}{1-\beta^2}\right) \\
(\tilde{\eta}_{0j}, \tilde{\eta}_{-1,j})' &\rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{(1+\lambda_1\lambda_2)\sigma_\eta^2}{(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} & \frac{(\lambda_1+\lambda_2)\sigma_\eta^2}{(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} \\ \frac{(\lambda_1+\lambda_2)\sigma_\eta^2}{(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} & \frac{(1+\lambda_1\lambda_2)\sigma_\eta^2}{(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} \end{pmatrix}\right) \\
&\rightsquigarrow N\left(\begin{pmatrix} X_{00} \\ X_{-1,0} \end{pmatrix}, \begin{pmatrix} X_{00} \\ X_{-1,0} \end{pmatrix}\right) \\
&\rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{(1+\lambda_1\lambda_2)\sigma_\eta^2}{(1-\beta^2)(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} & \frac{(\lambda_1+\lambda_2)\sigma_\eta^2}{(1-\beta^2)(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} \\ \frac{(\lambda_1+\lambda_2)\sigma_\eta^2}{(1-\beta^2)(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} & \frac{(1+\lambda_1\lambda_2)\sigma_\eta^2}{(1-\beta^2)(1-\lambda_1\lambda_2)(1-\lambda_1^2)(1-\lambda_2^2)} \end{pmatrix}\right).
\end{aligned} \tag{3.22}$$

Supongamos también, que la ecuación de observaciones tiene la misma forma dada por la ecuación (3.2):

$$Y_{ij} = m_{ij}X_{ij} + \epsilon_{ij}, \quad (i, j) \in R, \tag{3.23}$$

donde $\epsilon_{ij} \rightsquigarrow N(0, \sigma_\epsilon^2)$ son independientes entre sí y conjuntamente independientes con los η , $\tilde{\eta}$, X_{00} y $X_{-1,0}$, y los coeficientes m_{ij} son constantes conocidas. Así, la log-verosimilitud adoptará la expresión (salvo constantes)

$$\begin{aligned}
\log L &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij}x_{ij})^2 - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \log(\sigma_\eta^2) - \\
&\quad - \frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n [x_{ij} - (\lambda_1 + \lambda_2)x_{i-1,j} + \lambda_1\lambda_2x_{i-2,j} - \beta x_{i,j-1} + \\
&\quad + \beta(\lambda_1 + \lambda_2)x_{i-1,j-1} - \beta\lambda_1\lambda_2x_{i-2,j-1}]^2 - \frac{1}{2} \sum_{i=1}^m \log\left(\frac{\sigma_\eta^2}{1-\beta^2}\right) - \\
&\quad - \frac{1-\beta^2}{2\sigma_\eta^2} \sum_{i=1}^m [x_{i0} - (\lambda_1 + \lambda_2)x_{i-1,0} + \lambda_1\lambda_2x_{i-2,0}]^2 - \\
&\quad - \frac{1}{2} [2\log(\sigma_\eta^2) + \log\left(\frac{1}{(1-\beta^2)^2(1-\lambda_1\lambda_2)^2(1-\lambda_1^2)(1-\lambda_2^2)}\right)] - \\
&\quad - \frac{(1-\lambda_1\lambda_2)(1-\beta^2)}{2\sigma_\eta^2} [(1+\lambda_1\lambda_2)[x_{00}^2 + x_{-1,0}^2] - 2(\lambda_1 + \lambda_2)x_{00}x_{-1,0}] -
\end{aligned}$$

$$\begin{aligned}
 & -\frac{n}{2} \left[\log(\sigma_\eta^2) + \log\left(\frac{(1 + \lambda_1 \lambda_2)^2 - (\lambda_1 + \lambda_2)^2}{(1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)^2 (1 - \lambda_2^2)^2}\right) \right] - \\
 & \frac{(1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)^2 (1 - \lambda_2^2)^2}{2[(1 + \lambda_1 \lambda_2)^2 - (\lambda_1 + \lambda_2)^2] \sigma_\eta^4} \sum_{j=1}^n \left[\frac{(1 + \lambda_1 \lambda_2) \sigma_\eta^2}{(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \right. \\
 & \quad \left. [(x_{0j} - \beta x_{0,j-1})^2 + (x_{-1,j} - \beta x_{-1,j-1})^2] - \right. \\
 & \quad \left. -2 \frac{(\lambda_1 + \lambda_2) \sigma_\eta^2}{(1 - \lambda_1 \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} [(x_{0j} - \beta x_{0,j-1})(x_{-1,j} - \beta x_{-1,j-1})] \right].
 \end{aligned}$$

Si simplificamos factores comunes y agrupamos algunos términos obtenemos

$$\begin{aligned}
 \log L = & -\frac{nm}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij} x_{ij})^2 + \frac{m}{2} \log(1 - \beta^2) - \\
 & -\frac{nm + m + 2n + 2}{2} \log(\sigma_\eta^2) + \frac{n}{2} \log[(1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)(1 - \lambda_2^2)] + \\
 & + \frac{1}{2} \log[(1 - \beta^2)^2 (1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)(1 - \lambda_2^2)] - \\
 & -\frac{1}{2\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n [x_{ij} - (\lambda_1 + \lambda_2)x_{i-1,j} + \lambda_1 \lambda_2 x_{i-2,j} - \beta x_{i,j-1} + \\
 & + \beta(\lambda_1 + \lambda_2)x_{i-1,j-1} - \beta \lambda_1 \lambda_2 x_{i-2,j-1}]^2 - \\
 & -\frac{1 - \beta^2}{2\sigma_\eta^2} \sum_{i=1}^m [x_{i0} - (\lambda_1 + \lambda_2)x_{i-1,0} + \lambda_1 \lambda_2 x_{i-2,0}]^2 - \\
 & -\frac{(1 - \lambda_1 \lambda_2)(1 - \beta^2)}{2\sigma_\eta^2} [(1 + \lambda_1 \lambda_2)[x_{00}^2 + x_{-1,0}^2] - 2(\lambda_1 + \lambda_2)x_{00}x_{-1,0}] - \\
 & -\frac{(1 - \lambda_1 \lambda_2)}{2\sigma_\eta^2} \sum_{j=1}^n [(1 + \lambda_1 \lambda_2)[(x_{0j} - \beta x_{0,j-1})^2 + (x_{-1,j} - \beta x_{-1,j-1})^2] - \\
 & -2(\lambda_1 + \lambda_2)(x_{0j} - \beta x_{0,j-1})(x_{-1,j} - \beta x_{-1,j-1})].
 \end{aligned}$$

Ahora desarrollamos las potencias y los productos que aparecen en la expresión anterior, agrupando las sumatorias en factores comunes a los parámetros por estimar, obteniendo:

$$\begin{aligned}
 \log L = & -\frac{nm}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij} x_{ij})^2 + \frac{m}{2} \log(1 - \beta^2) - \\
 & -\frac{nm + m + 2n + 2}{2} \log(\sigma_\eta^2) + \frac{n}{2} \log[(1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)(1 - \lambda_2^2)] +
 \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \log[(1 - \beta^2)^2(1 - \lambda_1\lambda_2)^2(1 - \lambda_1^2)(1 - \lambda_2^2)] - \\
& - \frac{1}{2} [AA + (\lambda_1 + \lambda_2)^2 AB + \lambda_1^2 \lambda_2^2 AC + \beta^2 AD + \beta^2 (\lambda_1 + \lambda_2)^2 AE + \\
& + \beta^2 \lambda_1^2 \lambda_2^2 AF - (\lambda_1 + \lambda_2) BA + \lambda_1 \lambda_2 BB - \beta BC + \\
& + \beta (\lambda_1 + \lambda_2) BD - \beta \lambda_1 \lambda_2 BE - \lambda_1 \lambda_2 (\lambda_1 + \lambda_2) BF - \\
& - \beta (\lambda_1 + \lambda_2)^2 BG + \beta \lambda_1 \lambda_2 (\lambda_1 + \lambda_2) BH - \beta \lambda_1^2 \lambda_2^2 BI - \\
& - \beta^2 (\lambda_1 + \lambda_2) BJ + \beta^2 \lambda_1 \lambda_2 BK - \beta^2 \lambda_1 \lambda_2 (\lambda_1 + \lambda_2) BL],
\end{aligned}
\tag{3.24}$$

donde

$$AA = \sum_{i=-1}^m \sum_{j=0}^n x_{ij}^2$$

$$AB = \sum_{i=1}^m \sum_{j=0}^n x_{i-1,j}^2$$

$$AC = \sum_{i=3}^m \sum_{j=0}^n x_{i-2,j}^2$$

$$AD = \sum_{i=-1}^m \sum_{j=2}^n x_{i,j-1}^2$$

$$AE = \sum_{i=1}^m \sum_{j=2}^n x_{i-1,j-1}^2$$

$$AF = \sum_{i=3}^m \sum_{j=2}^n x_{i-2,j-1}^2$$

$$BA = 2 \sum_{i=0}^m \sum_{j=0}^n x_{i,j} x_{i-1,j}$$

$$BB = 2 \sum_{i=1}^m \sum_{j=0}^n x_{i,j} x_{i-2,j}$$

$$BC = 2 \sum_{i=-1}^m \sum_{j=1}^n x_{i,j} x_{i,j-1}$$

$$BD = 2 \sum_{i=0}^m \sum_{j=1}^n [x_{i,j} x_{i-1,j-1} + x_{i,j-1} x_{i-1,j}]$$

$$BE = 2 \sum_{i=1}^m \sum_{j=1}^n [x_{i,j} x_{i-2,j-1} + x_{i,j-1} x_{i-2,j}]$$

$$BF = 2 \sum_{i=2}^m \sum_{j=0}^n x_{i-1,j} x_{i-2,j}$$

$$BG = 2 \sum_{i=1}^m \sum_{j=1}^n x_{i-1,j} x_{i-1,j-1}$$

$$BH = 2 \sum_{i=1}^m \sum_{j=2}^n [x_{i-1,j} x_{i-2,j-1} + x_{i-2,j} x_{i-1,j-1}]$$

$$BI = 2 \sum_{i=3}^m \sum_{j=1}^n x_{i-2,j} x_{i-2,j-1}$$

$$BJ = 2 \sum_{i=0}^m \sum_{j=2}^n x_{i,j-1} x_{i-1,j-1}$$

$$BK = 2 \sum_{i=1}^m \sum_{j=2}^n x_{i,j-1} x_{i-1,j-1}$$

$$BL = 2 \sum_{i=2}^m \sum_{j=2}^n x_{i-1,j-1} x_{i-2,j-1}$$

Una vez que tenemos la expresión de la log-verosimilitud, para aplicar los algoritmos presentados anteriormente debemos calcular la función $Q(\psi/\psi^{(r)})$ que, como recordamos, no es más que la esperanza condicionada de la log-verosimilitud al conjunto de observaciones \mathbf{Y} , supuesto el valor estimado $\psi^{(r)} = (\lambda_1^{(r)}, \lambda_2^{(r)}, \beta^{(r)}, \sigma_\eta^{2(r)}, \sigma_\epsilon^{2(r)})$ en la r -ésima iteración. Esta función adopta ahora la expresión:

$$\begin{aligned} Q(\psi/\psi^{(r)}) = & -\frac{nm}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n [y_{ij}^2 + m_{ij}^2 [x_{ij}^{mn2} + P_{ij}^{mn}] - \\ & - m_{ij} y_{ij} x_{ij}^{mn}] + \frac{m}{2} \log(1 - \beta^2) - \\ & - \frac{nm + m + 2n + 2}{2} \log(\sigma_\eta^2) + \\ & + \frac{n}{2} \log[(1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2) (1 - \lambda_2^2)] + \\ & + \frac{1}{2} \log[(1 - \beta^2)^2 (1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2) (1 - \lambda_2^2)] - \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{2}[\overline{AA} + (\lambda_1 + \lambda_2)^2\overline{AB} + \lambda_1^2\lambda_2^2\overline{AC} + \beta^2\overline{AD} + \\
& + \beta^2(\lambda_1 + \lambda_2)^2\overline{AE} + \beta^2\lambda_1^2\lambda_2^2\overline{AF} - (\lambda_1 + \lambda_2)\overline{BA} + \\
& + \lambda_1\lambda_2\overline{BB} - \beta\overline{BC} + \beta(\lambda_1 + \lambda_2)\overline{BD} - \beta\lambda_1\lambda_2\overline{BE} - \\
& - \lambda_1\lambda_2(\lambda_1 + \lambda_2)\overline{BF} - \beta(\lambda_1 + \lambda_2)^2\overline{BG} + \\
& + \beta\lambda_1\lambda_2(\lambda_1 + \lambda_2)\overline{BH} - \beta\lambda_1^2\lambda_2^2\overline{BI} - \beta^2(\lambda_1 + \lambda_2)\overline{BJ} + \\
& + \beta^2\lambda_1\lambda_2\overline{BK} - \beta^2\lambda_1\lambda_2(\lambda_1 + \lambda_2)\overline{BL}],
\end{aligned} \tag{3.25}$$

donde

$$\begin{aligned}
\overline{AA} &= \sum_{i=-1}^m \sum_{j=0}^n [x_{ij}^{mn2} + P_{ij}^{mn}] \\
\overline{AB} &= \sum_{i=1}^m \sum_{j=0}^n [x_{i-1,j}^{mn2} + P_{i-1,j}^{mn}] \\
\overline{AC} &= \sum_{i=3}^m \sum_{j=0}^n [x_{i-2,j}^{mn2} + P_{i-2,j}^{mn}] \\
\overline{AD} &= \sum_{i=-1}^m \sum_{j=2}^n [x_{i,j-1}^{mn2} + P_{i,j-1}^{mn}] \\
\overline{AE} &= \sum_{i=1}^m \sum_{j=2}^n [x_{i-1,j-1}^{mn2} + P_{i-1,j-1}^{mn}] \\
\overline{AF} &= \sum_{i=3}^m \sum_{j=2}^n [x_{i-2,j-1}^{mn2} + P_{i-2,j-1}^{mn}] \\
\overline{BA} &= 2 \sum_{i=0}^m \sum_{j=0}^n [x_{i,j}^{mn} x_{i-1,j}^{mn} + P_{i,j,i-1,j}^{mn}] \\
\overline{BB} &= 2 \sum_{i=1}^m \sum_{j=0}^n [x_{i,j}^{mn} x_{i-2,j}^{mn} + P_{i,j,i-2,j}^{mn}] \\
\overline{BC} &= 2 \sum_{i=-1}^m \sum_{j=1}^n [x_{i,j}^{mn} x_{i,j-1}^{mn} + P_{i,j,i,j-1}^{mn}] \\
\overline{BD} &= 2 \sum_{i=0}^m \sum_{j=1}^n [x_{i,j}^{mn} x_{i-1,j-1}^{mn} + x_{i,j-1}^{mn} x_{i-1,j}^{mn} + P_{i,j,i-1,j-1}^{mn} + P_{i,j-1,i-1,j}^{mn}]
\end{aligned}$$

$$\overline{BE} = 2 \sum_{i=1}^m \sum_{j=1}^n [x_{i,j}^{mn} x_{i-2,j-1}^{mn} + x_{i,j-1}^{mn} x_{i-2,j}^{mn} + P_{i,j,i-2,j-1}^{mn} + P_{i,j-1,i-2,j}^{mn}]$$

$$\overline{BF} = 2 \sum_{i=2}^m \sum_{j=0}^n [x_{i-1,j}^{mn} x_{i-2,j}^{mn} + P_{i-1,j,i-2,j}^{mn}]$$

$$\overline{BG} = 2 \sum_{i=1}^m \sum_{j=1}^n [x_{i-1,j}^{mn} x_{i-1,j-1}^{mn} + P_{i-1,j,i-1,j-1}^{mn}]$$

$$\overline{BH} = 2 \sum_{i=1}^m \sum_{j=2}^n [x_{i-1,j}^{mn} x_{i-2,j-1}^{mn} + x_{i-2,j}^{mn} x_{i-1,j-1}^{mn} + P_{i-1,j,i-2,j-1}^{mn} + P_{i-2,j,i-1,j-1}^{mn}]$$

$$\overline{BI} = 2 \sum_{i=3}^m \sum_{j=1}^n [x_{i-2,j}^{mn} x_{i-2,j-1}^{mn} + P_{i-2,j,i-2,j-1}^{mn}]$$

$$\overline{BJ} = 2 \sum_{i=0}^m \sum_{j=2}^n [x_{i,j-1}^{mn} x_{i-1,j-1}^{mn} + P_{i,j-1,i-1,j-1}^{mn}]$$

$$\overline{BK} = 2 \sum_{i=1}^m \sum_{j=2}^n [x_{i,j-1}^{mn} x_{i-1,j-1}^{mn} + P_{i,j-1,i-1,j-1}^{mn}]$$

$$\overline{BL} = 2 \sum_{i=2}^m \sum_{j=2}^n [x_{i-1,j-1}^{mn} x_{i-2,j-1}^{mn} + P_{i-1,j-1,i-2,j-1}^{mn}]$$

siendo $x_{ij}^{mn} = E[x_{ij}/\mathbf{Y}]$, $P_{ij}^{mn} = \text{Var}[x_{ij}/\mathbf{Y}]$, y $P_{ijkl}^{mn} = \text{Cov}[x_{ij}, x_{kl}/\mathbf{Y}]$. Estas esperanzas y covarianzas condicionadas se calculan convenientemente construyendo una base ortogonal y proyectando sobre ella, de forma análoga a como se procedió en la sección 3.3. El paso M ahora depende tan sólo del algoritmo que elijamos.

La deducción del modelo $\text{AR}(1) \times \text{AR}(2)$ es, obviamente, simétrica a la del $\text{AR}(2) \times \text{AR}(1)$.

3.7 Modelo $\text{AR}(2) \times \text{AR}(2)$

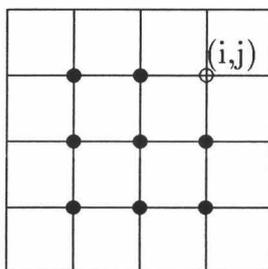
Consideramos en esta sección el caso de un proceso separable con estructura $\text{AR}(2)$ en ambas direcciones horizontal y vertical:

$$(1 - (\lambda_1 + \lambda_2)B_1 + \lambda_1\lambda_2B_1^2)(1 - (\beta_1 + \beta_2)B_2 + \beta_1\beta_2B_2^2)X_{ij} = \eta_{ij},$$

donde λ_1 , λ_2 , β_1 y β_2 tienen módulo menor que uno y η_{ij} es un ruido blanco gaussiano con varianza σ_η^2 . Desarrollando esta expresión obtenemos la siguiente

$$\begin{aligned} X_{ij} = & (\lambda_1 + \lambda_2)X_{i-1,j} - \lambda_1\lambda_2X_{i-2,j} + (\beta_1 + \beta_2)X_{i,j-1} - \\ & - (\lambda_1 + \lambda_2)(\beta_1 + \beta_2)X_{i-1,j-1} + \lambda_1\lambda_2(\beta_1 + \beta_2)X_{i-2,j-1} - \\ & - \beta_1\beta_2X_{i,j-2} + (\lambda_1 + \lambda_2)\beta_1\beta_2X_{i-1,j-2} - \lambda_1\lambda_2\beta_1\beta_2X_{i-2,j-2} + \eta_{ij}. \end{aligned} \quad (3.26)$$

Este modelo relaciona el valor en cada nodo con el valor en ocho nodos y el de un cierto ruido blanco, según se ilustra en el siguiente dibujo.



Desarrollando las condiciones iniciales, que en este caso están asociadas a dos filas y dos columnas adyacentes a los bordes Sur y Oeste, respectivamente, y siguiendo el mismo esquema que en la sección anterior para el modelo AR(2) × AR(1), obtenemos las relaciones siguientes:

$$(1 - \lambda_1 B_1)(1 - \lambda_2 B_1)X_{i0} = \tilde{\eta}_{i0}, \quad (3.27)$$

$$(1 - \lambda_1 B_1)(1 - \lambda_2 B_1)X_{i,-1} = \tilde{\eta}_{i,-1}, \quad (3.28)$$

$$(1 - \beta_1 B_2)(1 - \beta_2 B_2)X_{0j} = \tilde{\eta}_{0j}, \quad (3.29)$$

$$(1 - \beta_1 B_2)(1 - \beta_2 B_2)X_{-1,j} = \tilde{\eta}_{-1,j}, \quad (3.30)$$

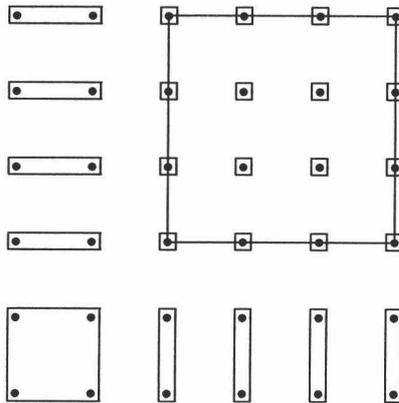
donde $(\tilde{\eta}_{i0}, \tilde{\eta}_{i,-1})'$ y $(\tilde{\eta}_{0j}, \tilde{\eta}_{-1,j})'$ tienen distribuciones normales bivariantes independientes entre sí y con todos los η_{ij} en la malla. Las medias de estas distribuciones son nulas y sus matrices de covarianzas vienen dadas por

$$\text{Var}(\tilde{\eta}_{i0}) = \frac{(1 + \beta_1\beta_2)\sigma_\eta^2}{(1 - \beta_1\beta_2)(1 - \beta_1^2)(1 - \beta_2^2)}$$

$$\begin{aligned} \text{Var}(\tilde{\eta}_{i,-1}) &= \frac{(1 + \beta_1\beta_2)\sigma_\eta^2}{(1 - \beta_1\beta_2)(1 - \beta_1^2)(1 - \beta_2^2)} \\ \text{Var}(\tilde{\eta}_{0j}) &= \frac{(1 + \lambda_1\lambda_2)\sigma_\eta^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \\ \text{Var}(\tilde{\eta}_{-1,j}) &= \frac{(1 + \lambda_1\lambda_2)\sigma_\eta^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \\ \text{Cov}(\tilde{\eta}_{i0}, \tilde{\eta}_{i,-1}) &= \frac{(\beta_1 + \beta_2)\sigma_\eta^2}{(1 - \beta_1\beta_2)(1 - \beta_1^2)(1 - \beta_2^2)} \\ \text{Cov}(\tilde{\eta}_{0j}, \tilde{\eta}_{-1,j}) &= \frac{(\lambda_1 + \lambda_2)\sigma_\eta^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)}. \end{aligned}$$

Además necesitaremos cuatro condiciones iniciales $(X_{00}, X_{-1,0}, X_{0,-1}, X_{-1,-1})'$ que serán independientes de los 'inputs' anteriores y se distribuirán conjuntamente según una normal multivariante con vector de medias nulo y matriz de covarianzas dada por la relación:

$$\begin{aligned} \text{Cov}(X_{ij}, X_{kl}) &= \\ &= \frac{[(1 - \lambda_2^2)\lambda_1^{|i-k|+1} - (1 - \lambda_1^2)\lambda_2^{|i-k|+1}][(1 - \beta_2^2)\beta_1^{|j-l|+1} - (1 - \beta_1^2)\beta_2^{|j-l|+1}]}{(1 - \lambda_1\lambda_2)(\lambda_1 - \lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)(1 - \beta_1\beta_2)(\beta_1 - \beta_2)(1 - \beta_1^2)(1 - \beta_2^2)} \sigma_\eta^2. \end{aligned} \tag{3.31}$$



Así, el modelo se puede formular, según ilustra el dibujo anterior, haciendo uso de la normalidad e independencia conjunta entre las variables $\{\eta_{ij} : i = 1, \dots, m, j = 1, \dots, n\}$ (cuadrados pequeños), $\{(\tilde{\eta}_{i0}, \tilde{\eta}_{i,-1})' : i = 1, \dots, m\}$

y $\{(\tilde{\eta}_{0j}, \tilde{\eta}_{-1,j})' : j = 1, \dots, n\}$ (rectángulos), y $(X_{00}, X_{-1,0}, X_{0,-1}, X_{-1,-1})'$ (cuadrado 2×2).

Supongamos, de nuevo, que la ecuación de observación tiene la misma forma dada por la ecuación (3.2):

$$Y_{ij} = m_{ij}X_{ij} + \epsilon_{ij}, \quad (i, j) \in R, \quad (3.32)$$

donde m_{ij} son constantes conocidas y $\epsilon_{ij} \rightsquigarrow N(0, \sigma_\epsilon^2)$ son independientes entre sí y conjuntamente independientes con los η , $\tilde{\eta}$, X_{00} , $X_{-1,0}$, $X_{0,-1}$ y $X_{-1,-1}$. Así, la log-verosimilitud adopta la siguiente expresión de la forma:

$$\begin{aligned} \log L = & -\frac{mn}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - m_{ij}x_{ij})^2 - \\ & -\frac{mn}{2} \log(\sigma_\eta^2) - \frac{1}{\sigma_\eta^2} \sum_{i=1}^m \sum_{j=1}^n [x_{ij} - (\lambda_1 + \lambda_2)x_{i-1,j} + \lambda_1\lambda_2x_{i-2,j} - \\ & -(\beta_1 - 1 + \beta_2)x_{i,j-1} + (\lambda_1 + \lambda_2)(\beta_1 + \beta_2)x_{i-1,j-1} - \lambda_1\lambda_2(\beta_1 + \beta_2)x_{i-2,j-1} + \\ & + \beta_1\beta_2x_{i,j-2} - (\lambda_1 + \lambda_2)\beta_1\beta_2x_{i-1,j-2} + \lambda_1\lambda_2\beta_1\beta_2x_{i-2,j-2}]^2 - \\ & -\frac{n}{2} \left[\log \left(\frac{(1 + \lambda_1\lambda_2)^2 - (\lambda_1 + \lambda_2)^2}{(1 - \lambda_1\lambda_2)^2(1 - \lambda_1^2)^2(1 - \lambda_2^2)^2} \right) + 2 \log(\sigma_\eta^2) \right] - \\ & -\frac{(1 - \lambda_1\lambda_2)^2(1 - \lambda_1^2)^2(1 - \lambda_2^2)^2}{2[(1 + \lambda_1\lambda_2)^2 - (\lambda_1 + \lambda_2)^2]\sigma_\eta^4} \sum_{j=1}^n \left[\frac{(1 + \lambda_1\lambda_2)\sigma_\eta^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} \right. \\ & \left. [(x_{0j} - (\beta_1 + \beta_2)x_{0,j-1} + \beta_1\beta_2x_{0,j-2})^2 + (x_{-1,j} - (\beta_1 + \beta_2)x_{-1,j-1} + \right. \\ & \left. + \beta_1\beta_2x_{-1,j-2})^2] - 2 \frac{(\lambda_1 + \lambda_2)\sigma_\eta^2}{(1 - \lambda_1\lambda_2)(1 - \lambda_1^2)(1 - \lambda_2^2)} [(x_{0j} - (\beta_1 + \beta_2)x_{0,j-1} + \right. \\ & \left. + \beta_1\beta_2x_{0,j-2})(x_{-1,j} - (\beta_1 + \beta_2)x_{-1,j-1} + \beta_1\beta_2x_{-1,j-2}) \right] - \\ & -\frac{m}{2} \left[\log \left(\frac{(1 + \beta_1\beta_2)^2 - (\beta_1 + \beta_2)^2}{(1 - \beta_1\beta_2)^2(1 - \beta_1^2)^2(1 - \beta_2^2)^2} \right) + 2 \log(\sigma_\eta^2) \right] - \\ & -\frac{(1 - \beta_1\beta_2)^2(1 - \beta_1^2)^2(1 - \beta_2^2)^2}{2[(1 + \beta_1\beta_2)^2 - (\beta_1 + \beta_2)^2]\sigma_\eta^4} \sum_{i=1}^m \left[\frac{(1 + \beta_1\beta_2)\sigma_\eta^2}{(1 - \beta_1\beta_2)(1 - \beta_1^2)(1 - \beta_2^2)} \right. \\ & \left. [(x_{i0} - (\lambda_1 + \lambda_2)x_{i-1,0} + \lambda_1\lambda_2x_{i-2,0})^2 + (x_{i,-1} - (\lambda_1 + \lambda_2)x_{i-1,-1} + \right. \\ & \left. + \lambda_1\lambda_2x_{i-2,-1})^2] - 2 \frac{(\beta_1 + \beta_2)\sigma_\eta^2}{(1 - \beta_1\beta_2)(1 - \beta_1^2)(1 - \beta_2^2)} [(x_{i0} - (\lambda_1 + \lambda_2)x_{i-1,0} + \right. \\ & \left. + \lambda_1\lambda_2x_{i-2,0})(x_{i,-1} - (\lambda_1 + \lambda_2)x_{i-1,-1} + \lambda_1\lambda_2x_{i-2,-1}) \right] - \end{aligned}$$

$$\begin{aligned}
 & -2 \log(\sigma_\eta^2) + 2 \log(1 - \beta_1 \beta_2) + 2 \log(1 - \beta_1^2) + 2 \log(1 - \beta_2^2) + \\
 & + 2 \log(1 - \lambda_1 \lambda_2) + 2 \log(1 - \lambda_1^2) + 2 \log(1 - \lambda_2^2) - \log(1 + \lambda_1 \lambda_2) - \\
 & - \frac{1}{2} \log \left((1 + \lambda_1 \lambda_2)^2 [(1 + \beta_1 \beta_2)^4 - 2(1 + \beta_1 \beta_2)^2 (\beta_1 + \beta_2)^2 + \right. \\
 & \left. + 3(\beta_1 + \beta_2)^4] + 2(1 + \beta_1 \beta_2)^2 (\lambda_1 + \lambda_2)^2 [2(\beta_1 + \beta_2)^2 - (1 + \beta_1 \beta_2)^2] \right) - \\
 & - \frac{A}{B} [(1 + \lambda_1 \lambda_2)(1 + \lambda_1 \lambda_2) [x_{00}^2 + x_{0,-1}^2 + x_{-1,0}^2 + x_{-1,-1}^2] + \\
 & + 2(1 + \lambda_1 \lambda_2)(\beta_1 + \beta_2) [x_{00} x_{0,-1} + x_{-1,0} x_{-1,-1}] + \\
 & + 2(1 + \beta_1 \beta_2)(\lambda_1 + \lambda_2) [x_{00} x_{-1,0} + x_{0,-1} x_{-1,-1}] + \\
 & + 2(\lambda_1 + \lambda_2)(\beta_1 + \beta_2) [x_{00} x_{-1,-1} + x_{0,-1} x_{-1,0}]],
 \end{aligned} \tag{3.33}$$

donde

$$\begin{aligned}
 A &= (1 + \lambda_1 \lambda_2 + \lambda_1 + \lambda_2)(1 + \lambda_1 \lambda_2 - \lambda_1 - \lambda_2)(1 + \beta_1 \beta_2 + \beta_1 + \beta_2) \\
 & \quad (1 + \beta_1 \beta_2 - \beta_1 - \beta_2) \\
 B &= (1 + \lambda_1 \lambda_2)^2 (1 - \lambda_1 \lambda_2)^2 (1 - \lambda_1^2)^2 (1 - \lambda_2^2)^2 (1 - \beta_1 \beta_2)^2 (1 - \beta_1^2)^2 (1 - \beta_2^2)^2 \\
 & \quad 2\sigma_\eta^8 [(1 + \lambda_1 \lambda_2)^2 [(1 + \beta_1 \beta_2)^4 - 2(1 + \beta_1 \beta_2)^2 (\beta_1 + \beta_2)^2 + 3(\beta_1 + \beta_2)^4] + \\
 & \quad + 2(1 + \beta_1 \beta_2)^2 (\lambda_1 + \lambda_2)^2 [2(\beta_1 + \beta_2)^2 - (1 + \beta_1 \beta_2)^2]].
 \end{aligned}$$

El resto del procedimiento seguiría los mismos pasos que en los casos anteriormente desarrollados.

3.8 Algunas consideraciones sobre otros modelos

3.8.1 Modelo AR(1) × MA(1)

En este caso estamos interesados en un modelo con estructura autorregresiva en dirección horizontal y de medias móviles en dirección vertical, ambas de orden uno y con raíces de sus respectivos polinomios de retardos con módulo mayor que uno. Dicho modelo se formula mediante la expresión:

$$X_{ij} = \lambda X_{i-1,j} + \eta_{ij} - \beta \eta_{i,j-1}, \tag{3.34}$$

donde $\eta_{ij} \rightsquigarrow N(0, \sigma_\eta^2)$ y son conjuntamente independientes. La función de autocovarianza puede escribirse como

$$\text{Cov}(X_{ij}, X_{kl}) = \begin{cases} \lambda^{|i-k|} \left(\frac{-\beta}{1+\beta^2} \right)^{|j-l|} \frac{(1+\beta^2)\sigma_\eta^2}{1-\lambda^2} & \text{si } |j-l| \leq 1 \\ 0 & \text{si } |j-l| > 1 \end{cases}$$

Ahora, igual que en los modelos que hemos estudiado anteriormente, interesaría obtener una expresión simple de la verosimilitud en forma de producto de verosimilitudes normales de dimensión reducida. Con este fin podemos expresar el modelo (3.34) de la forma siguiente:

$$(1 - \lambda B)X_{ij} = \tilde{\eta}_{ij},$$

siendo $\tilde{\eta}_{ij} = \eta_{ij} - \beta\eta_{i,j-1}$. Es decir, el modelo viene expresado en función de valores de ruido que son independientes en localizaciones con primeros subíndices distintos o segundos con diferencia mayor que uno. Así, la verosimilitud se factoriza en función de m normales multivariantes de dimensión n (los ruidos $\tilde{\eta}_i$ son conjuntamente independientes). Estas normales tienen toda matriz de covarianzas dada por

$$\text{Var}(\tilde{\eta}_{ij}) = (1 + \beta^2)\sigma_\eta^2 \text{Cov}(\tilde{\eta}_{ij}, \tilde{\eta}_{i,j-1}) = \text{Cov}(\tilde{\eta}_{ij}, \tilde{\eta}_{i,j+1}) = -\beta\sigma_\eta^2,$$

y el resto de las covarianzas nulas. En este caso, por tanto, la metodología aplicada en los casos anteriormente estudiados se beneficia de la condición de separabilidad sólo parcialmente, tanto más cuanto menor sea n .

3.8.2 Modelo MA(1) \times MA(1)

Supongamos ahora un modelo con estructura separable de medias móviles de orden uno en ambas direcciones. El modelo adopta la siguiente expresión:

$$X_{ij} = \eta_{ij} - \lambda\eta_{i-1,j} - \beta\eta_{i,j-1} + \lambda\beta\eta_{i-1,j-1} \quad (3.35)$$

Suponemos además que los coeficientes λ y β son menores que uno en valor absoluto. Puede verse fácilmente que la función de autocovarianzas de X viene dada por la siguiente expresión

$$\text{Cov}(X_{ij}, X_{kl}) = \left(\frac{-\lambda}{1 + \lambda^2} \right)^{|i-k|} \left(\frac{-\beta}{1 + \beta^2} \right)^{|j-l|} (1 + \lambda^2)(1 + \beta^2)\sigma_\eta^2,$$

para $|i - k| \leq 1$ y $|j - l| \leq 1$, siendo nula en otro caso. Razonando de forma análoga a como se hizo en el caso anterior, se observa que la función de verosimilitud no admite una factorización en términos de distribuciones de dimensión reducida. No obstante la matriz de covarianzas global contiene una proporción alta de ceros que aumenta con n y m .

3.8.3 Caso general no separable

Con respecto a los nodos de la malla (supuesto un esquema autorregresivo) el tratamiento en relación con la verosimilitud es análogo, pudiendo desarrollarla en función de los 'inputs' independientes η_{ij} . El problema surge en la modelización de las condiciones iniciales, ya que, en general, habrá que considerar su distribución conjunta (si bien ésta es conocida), no pudiendo reducirla como ocurre en el caso separable.

Capítulo 4

Aplicaciones mediante simulación

4.1 Introducción

En este capítulo presentamos una serie de simulaciones con el fin de mostrar cómo se desarrollaría en la práctica el procedimiento de estimación de los parámetros y alisamiento a partir de un conjunto de datos espaciales. Por simplicidad, hemos considerado el caso de un proceso espacial $AR(1) \times AR(1)$.

En primer lugar, para aplicar el procedimiento, debemos de disponer de datos espaciales. Estos se han generados mediante un programa que hemos llamado **SIMUDG** compilado en FORTRAN 77 y que usa rutinas de la librería IMSL, que se describirá en la próxima sección. Una vez que disponemos de los datos hemos de seleccionar el algoritmo que queremos aplicar (EM, ECM o MCECM) y estudiar si se pueden estimar adecuadamente los cuatro parámetros o solamente tres. De acuerdo con estos casos posibles, hemos desarrollado cinco programas **EMDG**, **EMDGE**, **ECMDG**, **ECMDGE** y **MCECMDG**. Los programas acabados en E hacen referencia a que se estima el valor de los cuatro parámetros, es decir, incluida la varianza de ϵ . En los dos primeros programas se utiliza el algoritmo EM y el máximo es buscado evaluando la función Q en una malla. El tercero y el cuarto lo hacen aplicando el algoritmo ECM utilizando las funciones g_s que se presentaron en el capítulo anterior. El último emplea el algoritmo MCECM realizando un paso E entre cada paso CM.

El objetivo de estas simulaciones es poner de manifiesto el buen comportamiento de estos algoritmos y los problemas de identificación que pueden surgir en ciertas situaciones. Así, cuando los valores $m_{i,j}$ son muy dispares entre sí,

se pueden estimar bien los cuatro parámetros conjuntamente, obteniéndose estimaciones muy próximas a los valores con que han sido generados los datos. También cuando los parámetros estructurales (λ y β) son 'altos' no se presenta el problema de identificación, ya que la estructura del proceso X es 'muy fuerte'. Si alguno es 'bajo', si fijamos σ_ϵ^2 obtenemos buenas estimaciones para los demás parámetros. También mostramos casos en los que no se dispone de todas las observaciones, comprobando que las estimaciones que obtenemos son cercanas a los valores de generación aunque la proporción de datos de que se disponga sea pequeña.

Por último se presentan también programas para el cálculo de la matriz de varianzas-covarianzas asintótica (V): **SECM**, que nos permite elegir la matriz DM^{ECM} , y **SECMDGP**, que calcula la matriz V .

4.2 Simulación de los datos

El programa **SIMUDG** genera observaciones de un proceso doblemente geométrico $AR(1) \times AR(1)$ (X) y además proporciona los datos resultantes de la ecuación de observación (Y) en una malla $M \times N$. En todos los casos estudiados en la sección 4.4 se ha tomado $M = N = 15$.

El programa permite al comienzo elegir entre las siguientes opciones:

- Tomar todas las constantes m_{ij} iguales a 1.
- Leer las constantes m_{ij} del fichero **SIMUDGM.DAT**, donde se suponen que están ordenadas lexicográficamente.
- Generar las constantes m_{ij} aleatoriamente de acuerdo a una distribución uniforme (A,B).

A continuación, el programa requiere que introduzcamos los valores de los parámetros propios del modelo que queremos simular, es decir, λ , β , σ_η^2 y σ_ϵ^2 . Por último, nos permite la omisión aleatoria de datos, demandándonos el número de datos a omitir si elegimos esta opción.

Los ficheros de salida de este programa son tres. En el fichero **DGXYMC.DAT**, aparecen en cada fila las localizaciones (i,j), y los valores de X , Y y m en la citada localización. En el fichero **DGYMC.DAT** aparecen todos los valores de Y y m ordenados lexicográficamente. En el fichero

DGYM.DAT aparecen los mismos valores que en el anterior pero en los nodos con datos omitidos aparecen con valores de Y y m iguales a cero (en el fichero anterior no se omite ningún dato). Este último fichero es el que utilizarán como entrada todos los programas de estimación y alisamiento.

Un organigrama de este programa se puede consultar en el apéndice C.

4.3 Estimación de los parámetros y alisamiento

Con el fin de aplicar los procedimientos descritos en el capítulo 3 hemos desarrollado varios programas todos ellos compilados en FORTRAN 77.

Los programas **EMDG** y **EMDGE** realizan la aplicación del algoritmo EM para procesos $AR(1) \times AR(1)$. El programa **EMDGE** utiliza el fichero de entrada DGYM.DAT que contiene los datos y_{ij} y el valor de las constantes m_{ij} . Como salida, proporciona aparecen los ficheros EMDG.PAR y EMDG.ALI. El primero muestra los valores estimados de los parámetros en cada iteración. En el segundo aparecen en orden lexicográfico las posiciones (i, j) , los datos observados y_{ij} y los datos alisados de X e indica la razón por la que se ha detenido el programa, si es por alcanzar convergencia o por haber llegado a la iteración máxima especificada. Al inicio de la ejecución el programa pide que introduzcamos los valores iniciales de los parámetros. Una vez introducidos éstos calcula los valores de alisamiento x_{ij}^{mn} y P_{ijkl}^{mn} . A partir de éstos obtiene los coeficientes que intervienen en la función Q y la maximiza en λ y β evaluándola en una malla regular con precisión especificada (por defecto una centésima). El valor de σ_η^2 se calcula en función de λ y β según se mostró en el capítulo 3, de la misma forma se obtiene σ_ϵ^2 en función de los valores alisados. En cada iteración se muestran en pantalla los valores $\lambda^{(p)}$, $\beta^{(p)}$, $\sigma_\eta^{2(p)}$ y $\sigma_\epsilon^{2(p)}$. Un organigrama de este programa aparece en el apéndice C.

Para el caso en que no se pueda estimar los cuatro parámetros conjuntamente el programa **EMDG** realiza el procedimiento de forma análoga considerando el valor σ_ϵ^2 fijo, igual a su valor inicial.

Para la aplicación del algoritmo ECM hemos desarrollado, bajo la misma idea anteriormente expuesta, dos programas **ECMDG** y **ECMDGE**. La estructura de estos programas es similar a la de los anteriores. La diferencia fundamental radica en que el paso de maximización de la función Q se realiza, para λ y β , de acuerdo con las funciones g_s referidas anteriormente, mediante la rutina IMSL/UVMID. El cálculo del máximo para σ_η^2 y σ_ϵ^2 se realiza de forma

análoga a como la realizan los programas descritos anteriormente. Los ficheros de salida de este programa son ECMDG.PAR, ECMDG.ALI y ECMDG.AIJ. El primero muestra los valores estimados en de los parámetros en cada iteración. En el segundo, al igual que en los dos programas anteriores, aparecen en orden lexicográfico las posiciones (i, j) , los datos observados y_{ij} y los datos alisados de X e indica la razón por la que se ha detenido el programa, si es por alcanzar convergencia o por haber llegado a la iteración máxima especificada. En el último aparecen el número total de iteraciones y los valores \bar{a}_{ij} que serán necesarios para calcular \mathbf{I}_{oc} .

También hemos aplicado el algoritmo MCECM a los datos de la simulación, para lo cual se ha desarrollado el programa MCECMDG. Su estructura, de nuevo es similar a la del programa ECMDG salvo que realiza alternativamente una maximización (condicionada) y un paso E.

Finalmente, hemos implementado el cálculo de la matriz de varianzas-covarianzas asintótica en el programa SECMDG. Este programa necesita tres ficheros de entrada. El primero, ECMM.OUT contiene los coeficientes \bar{a}_{ij} de la última iteración, que fueron almacenados en el fichero ECMDG.AIJ. El segundo, PARECM.OUT, contiene las estimaciones de máxima verosimilitud que resultan de la aplicación del algoritmo ECM, y que constituyen la última fila de ECMDG.PAR. El tercero, DMECM.DAT, contiene los valores de DM^{ECM} y es calculado mediante el programa SECM que necesita los ficheros de entrada: DGYM.DAT, que contiene los datos, el fichero ECMDG.PAR que contiene todos los valores de las iteraciones, y ECMDG.AIJ que indica el número total de iteraciones realizadas. La salida de este programa nos proporciona los valores $r_{ij}^{(p)}$ que al estabilizarse dan los valores a introducir en el fichero DMECM.DAT. Con estos ficheros, el programa SECMDGP calcula la matriz \mathbf{I}_{oc} , la matrix DM^{CM} y la matrix DM^{ECM} , que lee directamente, calculando a partir de éstas la matriz de varianzas-covarianzas asintótica, V .

4.4 Algunas simulaciones estudiadas

En esta sección se describen algunos de los múltiples casos estudiados. Se trata de mostrar con cada ejemplo algún aspecto concreto en relación con la aplicación de los métodos estudiados.

Realizamos una primera simulación sin ningún dato omitido y con los siguientes parámetros:

λ	β	σ_{η}^2	σ_{ϵ}^2
0.35	0.75	2.0	0.2

Los datos simulados aparecen en el apéndice A.

Como el valor de λ se puede considerar “bajo” (recordemos que la influencia de los valores del ruido η sobre el proceso X decae de forma exponencial) trataremos el problema de estimar tan sólo los parámetros $\psi^* = (\lambda, \beta, \sigma_{\eta}^2)$; de hecho, intentamos inicialmente estimar los cuatro parámetros conjuntamente pero el valor de $\sigma_{\epsilon}^{2(p)}$ decrecía rápidamente a cero, por lo que nos centramos en la estimación de los otros tres parámetros considerando que σ_{ϵ}^2 era conocido.

En primer lugar, se seleccionaron condiciones iniciales, $\psi^*(0)$, para comenzar el algoritmo EM. Para esto consideramos las observaciones y_{ij} ordenadas según el orden lexicográfico. Realizamos un análisis de estos datos como si fuesen una serie unidimensional, según se expone a continuación.

En el gráfico de la función de autocorrelación estimada observamos que las únicas correlaciones significativas son las correspondientes a los retardos 1 y primeros múltiplos de 15, éstos con decaimiento exponencial (ver figura 4.1). Un análisis completo pone de manifiesto que la serie responde a una estructura autorregresiva de orden 1 ordinaria y autorregresiva de orden 1 estacional con estacionariedad de orden 15.

Obtuvimos que el parámetro AR(1) ordinario es estimado con el valor 0.3366 y el parámetro AR(1) estacional con el valor 0.7394. El análisis de los residuos revela un comportamiento de acuerdo con la hipótesis de ruido blanco (ver figura 4.2), con varianza estimada 2.63601 (hay que hacer notar que esta varianza está afectada al mismo tiempo por las varianzas de los ruidos η y ϵ , por lo que es normal que sea sesgada positivamente con respecto a los otros dos valores). Como puede observarse, esta forma de proceder para la determinación de los valores iniciales para los parámetros parece bastante apropiada, ya que en cierta medida los valores que proporciona se encontrarán relativamente cerca de los valores propios del campo aleatorio observado.

Aplicamos el programa **EMDG** a los datos simulados asumiendo el valor $\sigma_{\epsilon}^2 = 0.2$ fijo, obteniendo como estimaciones de los parámetros las siguientes:

parámetro	valor estimado
λ	0.40
β	0.77
σ_{η}^2	2.08

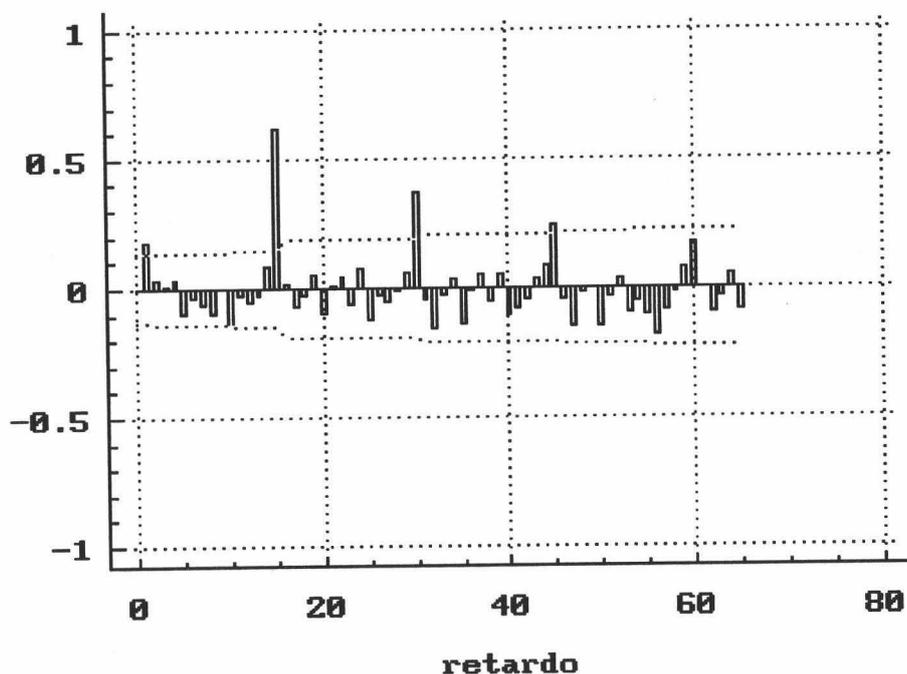


Figura 4.1: Función de autocorrelación estimada de los datos y_{ij} .

El alisamiento del campo aleatorio se puede observar en los gráficos de la columna que corresponde a la reconstrucción con el 100% de los datos observados (ver apéndice B). También se pueden observar que los errores para las predicciones son prácticamente iguales observándose que en los bordes se produce un incremento debido a que la información que proporciona la muestra tomada alrededor de estos puntos es menos completa.

También se ha hecho una estimación de los parámetros considerando un valor de $\sigma_\epsilon^2 = 0.7$ obteniendo las siguientes estimaciones de los parámetros:

parámetro	valor estimado
λ	0.45
β	0.80
σ_η^2	1.61

Aplicamos el programa **EMDG** a casos en que había *observaciones faltan-*

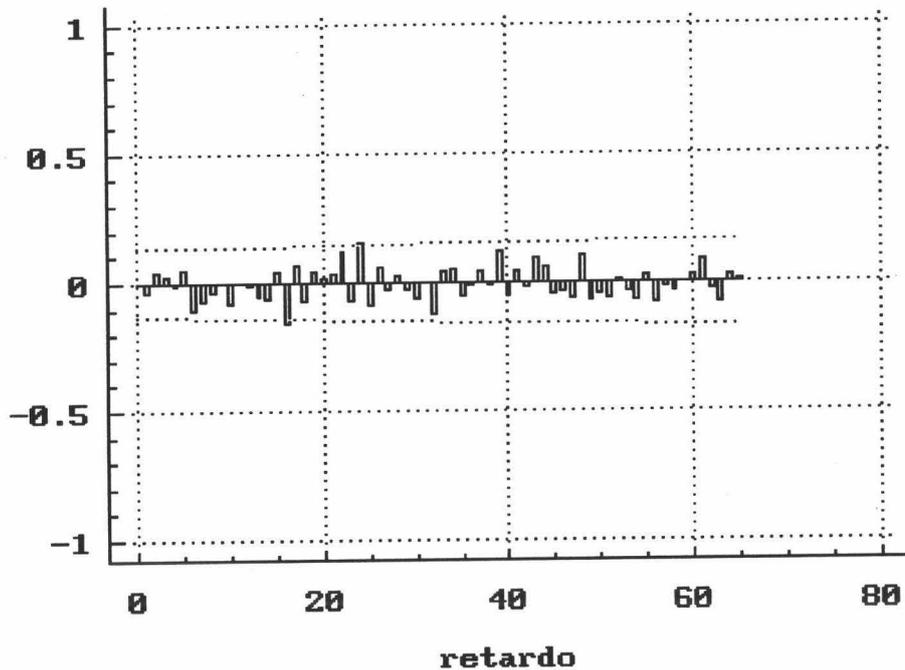


Figura 4.2: Función de autocorrelación estimada de residuos.

tes. Para ello simulamos tres casos seleccionando aleatoriamente los nodos a omitir con los mismos valores en los parámetros que en el caso anterior y la misma semilla para que los datos generados de X e Y (no omitidos) fueran los mismos. En el primer caso omitimos 56 datos, es decir los datos observados fueron prácticamente el 75% de las observaciones. En el segundo eliminamos 112 datos, observando así el 50%, y en el tercero eliminamos 169 datos. Las estimaciones de los parámetros obtenidas en estos casos aparecen en la siguiente tabla:

% de observaciones	λ	β	σ_{η}^2
100%	0.40	0.77	2.08
75%	0.41	0.75	2.20
50%	0.49	0.74	2.12
25%	0.54	0.79	1.71

Los valores alisados para estas estimaciones aparecen en las tres columnas

de gráficos de reconstrucciones anteriormente mencionados del apéndice B.

Aplicando el programa ECMDG a los datos de la simulación anterior, con un 100% de las observaciones, obtuvimos que el algoritmo necesitó, bajo las condiciones de estabilidad $|Q(\psi^* | \psi^{*(t)}) - Q(\psi^{*(t)} | \Psi^{*(t)})| < 10^{-6}$ y $\|\psi^{*(t+1)} - \psi^{*(t)}\| < 10^{-6}$, 12 iteraciones para estabilizarse, las mismas que el algoritmo EM necesitó (con 2 decimales de precisión para λ y β), obteniéndose como estimaciones de los parámetros

parámetro	valor estimado
λ	0.394637
β	0.769867
σ_η^2	2.080554

En los casos con observaciones omitidas obtuvimos

% de observaciones	λ	β	σ_η^2
100%	0.394637	0.7698677	2.080554
75%	0.411221	0.7548162	2.191500
50%	0.505747	0.7917352	1.988802
25%	0.510674	0.7955521	1.735017

Para el caso del 100% de observaciones se obtuvo que la matriz de varianzas-covarianzas de los estimadores asintótica era estimada por

$$\begin{pmatrix} 0.0049769 & 0.0001704 & -0.0021692 \\ 0.0001749 & 0.0022761 & -0.0029611 \\ -0.0007130 & -0.0016445 & 0.0502434 \end{pmatrix},$$

la cual se puede considerar aproximadamente simétrica. Con el 50% de las observaciones se obtuvo la matriz:

$$\begin{pmatrix} 0.0092586 & 0.0009376 & -0.0114871 \\ 0.0012445 & 0.0020861 & -0.0078834 \\ -0.0096494 & -0.0012107 & 0.0959227 \end{pmatrix}.$$

También aplicamos el programa MCECMDG a los datos de la simulación original, necesitando el algoritmo 12 iteraciones para estabilizarse con las mismas condiciones de parada, obteniéndose como estimación de los parámetros

parámetro	valor estimado
λ	0.394637
β	0.769867
σ_{η}^2	2.080554

Como conocemos en la simulación los valores de todos los datos x_{ij} podemos calcular la estimación de máxima verosimilitud de los parámetros, la cual nos puede servir de apoyo para comparar los distintos métodos, aunque en este caso todas las estimaciones proporcionadas son muy parecidas. Dichas estimaciones aparecen en la siguiente tabla

parámetro	valor estimado
λ	0.367581
β	0.758684
σ_{η}^2	2.13211

Se realizó otra simulación (los datos correspondientes aparecen también en el apéndice A) para los siguientes valores de los parámetros:

λ	β	σ_{η}^2	σ_{ϵ}^2
0.8	0.8	2.0	0.5

Las estimaciones usando los programas EMDGE y ECMDGE fueron

parámetro	EM	ECM
λ	0.81	0.815643
β	0.79	0.793610
σ_{η}^2	2.524174	2.484353
σ_{ϵ}^2	0.197088	0.211348

Otra simulación (véase apéndice A) con las constantes m_{ij} generadas aleatoriamente de acuerdo con una distribución uniforme (0,10) dio lugar a las siguientes estimaciones:

parámetro	valor generador	EM	ECM
λ	0.3	0.32	0.326030
β	0.8	0.80	0.801569
σ_{η}^2	3.0	3.159516	3.142186
σ_{ϵ}^2	0.7	0.669541	0.755964

aunque el valor de λ es relativamente bajo, debido a la heterogeneidad de los valores m_{ij} , es posible obtener una 'buena' estimación para los cuatro parámetros conjuntamente.

Hay que hacer notar, que en el estudio que hemos realizado, no hemos estudiado en ningún caso la "bondad" de los datos, es decir, si éstos tienen un comportamiento concordante con las hipótesis. De hecho, en un primer estudio no muy detallado apreciamos que, por ejemplo, los datos de la primera simulación no tenían media cero sino que esta era algo superior a dicho valor. Eliminamos esta media y obtuvimos las siguientes estimaciones usando el programa EMDG:

parámetro	valor estimado
λ	0.38
β	0.75
σ_{η}^2	2.08

que se encuentran más próximas a las estimaciones obtenidas por máxima verosimilitud.

La extensión a modelos más complejos no es muy complicada con algunas salvedades. El paso E, común a todos los algoritmos presentados, tendrá su estructura propia, dependiendo claramente del modelo que se intente ajustar, así por ejemplo, para el modelo $AR(2) \times AR(1)$ deberemos de calcular las constantes $\overline{AA}, \dots, \overline{AF}, \overline{BA}, \dots, \overline{BL}$. Una vez que se dispone de la función Q , sugerimos adaptar el algoritmo ECM o implementar un método para la búsqueda del máximo de funciones multiparamétricas, ya que la evaluación de la función Q aumenta de forma exponencial a medida que aumenta el número de parámetros a estimar. Además, la adaptación del algoritmo ECM solamente consiste en introducir tantos pasos de maximización unidimensional como parámetros nuevos aparezcan en la función Q .

Capítulo 5

Enfoques alternativos y problemas abiertos

5.1 Algunos enfoques alternativos

En el capítulo anterior hemos considerado que el proceso espacial de interés tenía parámetro discreto y era observado en nodos situados sobre una malla. Sin embargo, hay en la práctica multitud de situaciones, por ejemplo cuando se observan procesos continuos en localizaciones irregularmente distribuidas en el espacio, en que los datos no responden a una disposición tal que se le pueda ajustar una malla. Una solución posible a este problema podría ser, a la vista de los datos, definir alguna malla de forma que las observaciones estén localizadas satisfactoriamente “próximas” a algún nodo (eventualmente, podría interesar mallas no rectangulares). Ello requiere considerar el problema, en absoluto trivial, de la aproximación de procesos de parámetro continuo mediante modelos de parámetro discreto. Un enfoque alternativo natural a esta posible solución consiste en abordar el problema directamente, manteniendo la definición del proceso (por ejemplo, en términos de ecuaciones en derivadas parciales estocásticas) y localizaciones de observación originales, tratando de aplicar alguna metodología adecuada al caso. En el caso de procesos estacionarios de segundo orden, el problema puede enfocarse convenientemente haciendo uso de la relación que liga la función de autocovarianzas y el espectro en términos de la Transformada de Fourier (Teorema de Wiener–Khintchine) (ver por ejemplo, los trabajos de Whittle (1954), Jones (1979), Vecchia (1988), etc.). Una dificultad importante de esta aproximación es que no siempre es fácil o

posible disponer de una expresión analítica para la función de autocovarianzas (ver por ejemplo, Whittle (1954), Heine (1955), etc.), obteniendo como solución de la correspondiente integral sobre el espectro. Angulo et al. (1992,1994), proponen hacer una aproximación de dicha integral en términos de la transformada de Fourier discreta, lo que permite obtener una aproximación numérica de la solución sin necesidad de disponer de su forma analítica. Alonso et al. (1993) prueban que los estimadores de los parámetros involucrados en el modelo haciendo uso de esta aproximación convergen a los estimadores que se obtendrían sin el uso de la misma. Los principales resultados sobre este aspecto se exponen a continuación

Sea $z(x, y)$ un proceso espacial estacionario con estructura dada por la ecuación en derivadas parciales estocástica de segundo orden

$$a_{20} \frac{\partial^2 z(x, y)}{\partial x^2} + a_{02} \frac{\partial^2 z(x, y)}{\partial y^2} + 2a_{11} \frac{\partial^2 z(x, y)}{\partial x \partial y} - z(x, y) = \epsilon(x, y) \quad (5.1)$$

donde $(x, y) \in \mathbb{R}^2$, $\epsilon(x, y)$ es un ruido blanco gaussiano con media cero y varianza σ_ϵ^2 , y los coeficientes a_{20} , a_{02} y a_{11} son constantes.

Las funciones de autocovarianza y autocorrelación de z están definidas como

$$C(\tau_x, \tau_y) = E[z(x, y), z(x', y')], \quad \tau_x = x' - x, \quad \tau_y = y' - y$$

$$R(\tau_x, \tau_y) = \frac{C(\tau_x, \tau_y)}{C(0, 0)}, \quad C(0, 0) = \sigma_z^2$$

Se asume que la observación de z en un punto genérico (x, y) conlleva algún error de medida aditivo $\eta(x, y)$ con estructura de ruido blanco gaussiano con media cero y varianza σ_η^2 . Denotamos por $\mathbf{z}^* = (z_1^*, z_2^*, \dots, z_n^*)'$, el vector de observaciones asociado al conjunto de n localizaciones diferentes $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, donde cada componente representa

$$z_k^* = z_k + \eta_k$$

con $z_k^* = z^*(x_k, y_k)$, $z_k = z(x_k, y_k)$ y $\eta_k = \eta(x_k, y_k)$, para $k = 1, 2, \dots, n$.

Consideramos el problema de estimación relativo a los parámetros y coeficientes que aparecen en el modelo (5.1), a_{20} , a_{02} , a_{11} , σ_ϵ^2 y σ_η^2 (o, alternativamente, reemplazando σ_ϵ^2 por σ_z^2).

La función de verosimilitud asociada está dada por la expresión

$$L(\theta) = \frac{1}{(2\pi)^{n/2} |C^*|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{z}^{*'} C^{*-1} \mathbf{z}^*\right\} \quad (5.2)$$

donde θ representa el vector de parámetros $\theta = (a_{20}, a_{02}, a_{11}, \sigma_\epsilon^2, \sigma_\eta^2)$, y C^* es la matriz de varianzas-covarianzas de \mathbf{z}^* ,

$$C^* = C + \sigma_\eta^2 I$$

donde C es la matriz de varianzas-covarianzas de \mathbf{z} e I la matriz identidad de orden n .

Asumimos, en principio, que los elementos de C^* pueden calcularse cuando conocemos los valores de los parámetros, θ (se ilustrará esta suposición posteriormente). Consideramos el problema de estimación máximo verosímil:

$$\max L(a_{20}, a_{02}, a_{11}, \sigma_\epsilon^2, \sigma_\eta^2)$$

Para obtener una solución, algunas manipulaciones son de ayuda (ver Jones (1989)). Primero calculamos $-2\ln L$,

$$l(a_{20}, a_{02}, a_{11}, \sigma_\epsilon^2, \sigma_\eta^2) = \ln |C^*| + \mathbf{z}^{*'} C^{*-1} \mathbf{z}^*$$

(los términos constantes han sido eliminados); la función ahora deberá de ser minimizada.

A continuación, l puede reformularse con respecto a σ_z^2 (implícitamente, σ_ϵ^2), como sigue. Definimos D como

$$C^* = \sigma_z^2 D, \quad D = R + sI$$

donde R es la matriz de correlación de z para los puntos observados, y $s = \sigma_\eta^2 / \sigma_z^2$. l puede escribirse como

$$l = n \ln \sigma_z^2 + \ln |D|^{1/2} + \frac{1}{\sigma_z^2} \mathbf{z}^{*'} D^{-1} \mathbf{z}^*$$

Derivando l con respecto a σ_z^2 e igualando a cero, obtenemos

$$\sigma_z^2 = \frac{1}{n} \mathbf{z}^{*'} D^{-1} \mathbf{z}^*$$

Sustituyendo esta expresión en l y eliminando los términos constantes,

$$l'(a_{20}, a_{02}, a_{11}, s) = \ln D + n \ln \mathbf{z}' D^{-1} \mathbf{z}^*$$

Finalmente, el cálculo de l' puede simplificarse mediante una descomposición de Cholesky de D ,

$$D = U'U, \quad U \text{ matriz triangular superior,}$$

llegando a las expresiones

$$l' = n \ln \sum_{i=1}^n a_i^2 + 2 \sum_{i=1}^n \ln u_{ii}$$

$$\sigma_z^2 = \frac{1}{n} \sum_{i=1}^n a_i^2,$$

donde a_i es la componente i -ésima del vector

$$\mathbf{a} = (U')^{-1} \mathbf{z}^*.$$

De (5.1) y usando el teorema de Wiener-Khinchine, obtenemos la siguiente representación espectral para la función de covarianza de z ,

$$C(\tau_x, \tau_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\sigma_\epsilon^2 e^{i(\lambda_x \tau_x + \lambda_y \tau_y)}}{[a_{20} \lambda_x^2 + a_{02} \lambda_y^2 + 2a_{11} \lambda_x \lambda_y + 1]^2} d\lambda_x d\lambda_y$$

Usando la definición de integral impropia y un cambio conveniente en las variables, obtenemos la aproximación

$$C_M(\tau_x, \tau_y) = \frac{1}{4M^2} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \frac{\sigma_\epsilon^2 e^{i2M(\omega_x \tau_x + \omega_y \tau_y)}}{[a_{20} \omega_x^2 + a_{02} \omega_y^2 + 2a_{11} \omega_x \omega_y + \frac{1}{4M^2}]^2} d\omega_x d\omega_y$$

La doble integral puede aproximarse mediante sumas de Riemann, y podemos escribir finalmente la siguiente aproximación numérica para $C(\tau_x, \tau_y)$ (ver Angulo et al. (1992)):

$$C_D(\tau_x, \tau_y) = \frac{1}{4M^2 N_x N_y} \sum_{p=0}^{N_x-1} \sum_{q=0}^{N_y-1} \frac{\sigma_\epsilon^2 e^{i2M(\omega_p \tau_x + \omega_q \tau_y)}}{[a_{20} \omega_p^2 + a_{02} \omega_q^2 + 2a_{11} \omega_p \omega_q + \frac{1}{4M^2}]^2} d\omega_x d\omega_y$$

con $\omega_p = p/N_x - 1/2$ y $\omega_q = q/N_y - 1/2$ (implícitamente, se está considerando una transformada de Fourier discreta). Por tanto, los elementos de la matriz de correlación R , que interviene en el cálculo de l' , pueden obtenerse mediante la expresión

$$R(\tau_x, \tau_y) = \sum_{p=0}^{N_x-1} \sum_{q=0}^{N_y-1} W_{pq} \cos[2M(\omega_p \tau_x + \omega_q \tau_y)]$$

donde

$$W_{pq} = \frac{\Delta_{pq}}{\Delta}$$

$$\Delta_{pq} = \left[a_{20}\omega_p^2 + a_{02}\omega_q^2 + 2a_{11}\omega_p\omega_q + \frac{1}{4M^2} \right]^{-2}$$

$$\Delta = \sum_{p=0}^{N_x-1} \sum_{q=0}^{N_y-1} \Delta_{pq},$$

con ω_p y ω_q definidas como antes.

A continuación, probamos que la aproximación numérica anteriormente propuesta de las covarianzas de \mathbf{z} es adecuada con respecto al problema de estimación por máxima verosimilitud, mostrando su comportamiento asintótico en términos de convergencia puntual.

N representará ahora los tres valores (M, N_x, N_y) conjuntamente, y por ' $N \rightarrow \infty$ ' denotaremos la condición conjunta ' $M \rightarrow \infty, N_x \rightarrow \infty$ y $N_y \rightarrow \infty$ ' (en el orden apropiado).

Sabemos que dados τ_x y τ_y ,

$$C_D(\tau_x, \tau_y) \rightarrow C(\tau_x, \tau_y), \text{ cuando } N \rightarrow \infty.$$

Por tanto

$$C_D^* \rightarrow C^*, \text{ cuando } N \rightarrow \infty,$$

donde $C_D^* = C_D + \sigma_\eta^2 I$ y C_D es la matriz de covarianzas obtenida mediante la aproximación.

Debido a que C^* es definida positiva, podemos afirmar que para cada \mathbf{z}^*

$$L_D(\theta, \mathbf{z}^*) \rightarrow L(\theta, \mathbf{z}^*), \text{ cuando } N \rightarrow \infty,$$

donde L_D es la función de verosimilitud L reemplazando C^* por C_D^* .

Sea $\hat{\theta}_D$ una solución al problema $\max L_D(\theta)$ y $\hat{\theta}$ una solución al problema $\max L(\theta)$. Por simplicidad, asumamos que $\hat{\theta}_D$ se busca de entre los elementos de un conjunto finito G (por ejemplo, la intersección de un hipercubo y una malla regular); en realidad esta suposición no es restrictiva, ya que computacionalmente la precisión nos hace buscar los valores estimados en una malla regular y podemos exigir que el conjunto sea tan grande como queramos para asegurarnos de que los valores de la solución están contenidos en dicho conjunto.

Para \mathbf{z}^* fijo, sea $\hat{\theta}_D(\mathbf{z}^*)$ una solución del problema

$$\max L_D(\theta, \mathbf{z}^*).$$

(Esta solución depende, obviamente, de la elección de N).

Sea S el conjunto de soluciones del problema

$$\max L(\theta, \mathbf{z}^*)$$

Teorema 5.1.3 (Alonso et al. (1993), p. 21)

Bajo las condiciones y notación anteriores, se verifica que

$$d(\hat{\theta}_D(\mathbf{z}^*), S) \rightarrow 0, \text{ cuando } N \rightarrow \infty,$$

donde $d(\cdot, \cdot)$ es la distancia euclídea en \mathbb{R}^n .

Demostración.

Supongamos que la tesis del teorema no es cierta. Entonces, por la compacidad de G , existe una función $\sigma : \mathbb{N} \rightarrow \mathbb{R}^3$ que define la sucesión

$$\{\theta_{\sigma(n)}(\mathbf{z}^*)\}_{n \in \mathbb{N}},$$

con $\sigma(n) \rightarrow \infty$, cuando $n \rightarrow \infty$, que converge a un punto $\theta_1 \in G \setminus S$.

Por la definición de $\hat{\theta}_{\sigma(n)}(\mathbf{z}^*)$ sabemos que

$$L_{\sigma(n)} \hat{\theta}_{\sigma(n)}(\mathbf{z}^*), \mathbf{z}^* \geq L_{\sigma(n)}(\theta_0),$$

para cualquier $\theta_0 \in S$, ($S \neq \emptyset$).

El miembro de la desigualdad de la derecha converge a $L(\theta_0)$, cuando $n \rightarrow \infty$. Sin embargo, el miembro de la izquierda lo hace a $L(\theta_1)$, cuando $n \rightarrow \infty$,

debido a la convergencia uniforme de $L_{\sigma(n)}$ sobre G (observar que $L_{\sigma(n)}$ es continua y G es finito).

Por tanto, obtenemos que

$$L(\theta_1) \geq L(\theta_0),$$

y esto implica que $\theta_1 \in S$ al pertenecer θ_0 , lo cual contradice la hipótesis de que $\theta_1 \in G \setminus S$. ■

5.2 Otros problemas abiertos

En esta última sección de la memoria se mencionan algunos problemas no abordados en la misma que se han planteado durante su realización y que son actualmente objeto de nuestra investigación.

Desde el punto de vista de la modelización tiene interés el problema de reversibilidad del campo aleatorio. Si estamos interesados en estimar, por ejemplo, un proceso $AR(1) \times AR(1)$ ¿influye de algún modo la dirección en que se diseñe el campo?, es decir, ¿son equivalentes los siguientes modelos?:

$$X_{ij} = \lambda X_{i-1,j} + \beta X_{i,j-1} - \lambda\beta X_{i-1,j-1} + \eta_{ij}$$

$$X_{ij} = \lambda X_{i+1,j} + \beta X_{i,j-1} - \lambda\beta X_{i+1,j-1} + \eta_{ij}^*$$

$$X_{ij} = \lambda X_{i-1,j} + \beta X_{i,j+1} - \lambda\beta X_{i-1,j+1} + \eta_{ij}^{**}$$

$$X_{ij} = \lambda X_{i+1,j} + \beta X_{i,j+1} - \lambda\beta X_{i+1,j+1} + \eta_{ij}^{***}.$$

La respuesta es afirmativa, si bien los ruidos no son independientes entre sí y su estructura de covarianza viene dada por las expresiones siguientes:

$$\text{Var}(\eta_{ij}) = \text{Var}(\eta_{ij}^*) = \text{Var}(\eta_{ij}^{**}) = \text{Var}(\eta_{ij}^{***}) = \sigma_\eta^2$$

$$\text{Cov}(\eta_{i-k,j-p}, \eta_{ij}^*) = \begin{cases} \lambda^k(1-\lambda^2)\sigma_\eta^2 & k \geq 0, p = 0 \\ -\lambda\sigma_\eta^2 & k = -1, p = 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$\text{Cov}(\eta_{i-k,j-p}, \eta_{ij}^{**}) = \begin{cases} \beta^p(1-\beta^2)\sigma_\eta^2 & k = 0, p \geq 0 \\ -\beta\sigma_\eta^2 & k = 0, p = -1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\begin{aligned}
 \text{Cov}(\eta_{i-k,j-p}, \eta_{ij}^{***}) &= \begin{cases} \lambda^k \beta^p (1 - \lambda^2)(1 - \beta^2) \sigma_\eta^2 & k \geq 0, p \geq 0 \\ -\lambda^k \beta (1 - \lambda^2) \sigma_\eta^2 & k \geq 0, p = -1 \\ -\lambda \beta^p (1 - \beta^2) \sigma_\eta^2 & k = -1, p \geq 0 \\ \lambda \beta \sigma_\eta^2 & k = -1, p = -1 \\ 0 & \text{en otro caso} \end{cases} \\
 \text{Cov}(\eta_{i-k,j-p}^*, \eta_{ij}^{***}) &= \begin{cases} \beta^p (1 - \beta^2) \sigma_\eta^2 & k = 0, p \geq 0 \\ -\beta \sigma_\eta^2 & k = 0, p = -1 \\ 0 & \text{en otro caso} \end{cases} \\
 \text{Cov}(\eta_{i-k,j-p}^*, \eta_{ij}^{**}) &= \begin{cases} \lambda^{-k} \beta^p (1 - \lambda^2)(1 - \beta^2) \sigma_\eta^2 & k \leq 0, p \geq 0 \\ -\lambda^{-k} \beta (1 - \lambda^2) \sigma_\eta^2 & k \leq 0, p = -1 \\ -\lambda \beta^p (1 - \beta^2) \sigma_\eta^2 & k = 1, p \geq 0 \\ \lambda \beta \sigma_\eta^2 & k = 1, p = -1 \\ 0 & \text{en otro caso} \end{cases} \\
 \text{Cov}(\eta_{i-k,j-p}^{**}, \eta_{ij}^{***}) &= \begin{cases} \lambda^k (1 - \lambda^2) \sigma_\eta^2 & k \geq 0, p = 0 \\ -\lambda \sigma_\eta^2 & k = -1, p = 0 \\ 0 & \text{en otro caso} \end{cases}
 \end{aligned}$$

Otros problemas de especial importancia se citan a continuación:

- Diseño óptimo de una malla de ajuste a observaciones localizadas irregularmente.
- Estudio de modelos con estructura más compleja (por ejemplo, ARMA separables y no separables, modelos de regresión espacial, modelos no estacionarios, etc.).
- Extensión al caso en que la variable de interés sea multivariante.
- Desarrollo de una metodología para la elaboración de modelos, que incluya los aspectos de identificación, estimación y validación (tipo Box-Jenkins).
- Estudio de modelos no gaussianos, modelos con variables discretas, modelos con variables cualitativas, etc..
- Estudio de aspectos relativos al problema de aproximación de modelos continuos mediante modelos discretos.

Apéndice A

Datos simulados

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.35$, $\beta = 0.75$ y $\sigma_\eta^2 = 2$, datos Y_{ij} de la ecuación de observación con todos los $m_{ij} = 1$ y $\sigma_\epsilon^2 = 0.2$, datos de X alisados y errores esperados.

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1 1	3.565483	4.062298	3.788110	0.183585
2 1	1.341963	1.918847	1.971350	0.181651
3 1	2.322185	2.305162	2.151526	0.181646
4 1	-1.347617	-1.227081	-1.048239	0.181646
5 1	-2.701298	-3.426069	-3.240749	0.181646
6 1	-3.519091	-3.701221	-3.613933	0.181646
7 1	-5.418998	-4.918092	-4.765337	0.181646
8 1	-2.056990	-1.903632	-1.642877	0.181646
9 1	2.376616	2.316155	2.147302	0.181646
10 1	0.330162	-0.563863	-0.553964	0.181646
11 1	-4.606561	-3.751174	-3.605839	0.181646
12 1	0.978110	0.691155	0.724647	0.181646
13 1	-0.118302	0.271384	0.079180	0.181646
14 1	-1.212180	-1.721247	-1.602261	0.181651
15 1	0.500941	0.531147	0.587911	0.183585
1 2	1.106009	0.535414	0.746855	0.175757
2 2	1.582366	1.324249	1.379038	0.173220
3 2	2.384358	1.923487	1.770206	0.173208
4 2	1.457827	2.011415	2.033755	0.173208
5 2	0.389788	0.890241	0.627125	0.173208
6 2	-0.585789	-0.261146	-0.316913	0.173208
7 2	-0.380235	-1.195831	-1.277755	0.173208
8 2	2.650302	2.892981	2.578045	0.173208
9 2	2.328302	2.638155	2.743985	0.173208
10 2	0.870360	0.063864	0.136081	0.173208
11 2	-2.310778	-2.557553	-2.491177	0.173208
12 2	1.946970	1.606965	1.538189	0.173208
13 2	-1.160373	-1.816912	-1.487335	0.173208
14 2	-0.625260	-0.721370	-0.804731	0.173220
15 2	2.424585	2.023899	2.040035	0.175757
1 3	0.254407	0.928657	1.005478	0.175717
2 3	1.462085	2.024453	1.869127	0.173159
3 3	-0.886270	-1.007741	-0.722358	0.173147
4 3	1.974845	3.280179	2.964023	0.173147
5 3	-0.739446	-0.682381	-0.627991	0.173147
6 3	-0.953320	-0.914847	-0.854642	0.173147
7 3	-2.035827	-1.877644	-1.887997	0.173147
8 3	1.961053	1.892569	1.923950	0.173147
9 3	3.484118	3.474646	3.513544	0.173147
10 3	2.604234	2.438938	2.271719	0.173147

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
11 3	0.230612	0.084809	0.196561	0.173147
12 3	3.303419	3.950325	3.839985	0.173147
13 3	1.035067	1.579562	1.429635	0.173147
14 3	0.399183	0.731504	0.733885	0.173159
15 3	4.460101	4.205260	4.168635	0.175717
1 4	1.721876	2.002561	2.009926	0.175716
2 4	1.742069	1.735557	1.587633	0.173158
3 4	-1.712305	-1.522678	-1.310791	0.173147
4 4	1.135263	0.880484	1.035333	0.173147
5 4	-1.765706	-2.122092	-2.094833	0.173147
6 4	-0.858743	-0.530987	-0.712411	0.173147
7 4	-1.597033	-2.488470	-2.348437	0.173147
8 4	1.422393	1.842382	1.823248	0.173147
9 4	4.818954	4.404212	4.407749	0.173147
10 4	2.257835	2.471390	2.491299	0.173147
11 4	2.513007	2.791934	2.647126	0.173147
12 4	3.991704	4.082114	3.980403	0.173147
13 4	2.124664	1.293349	1.385177	0.173147
14 4	0.856867	0.925426	0.933478	0.173158
15 4	5.141075	5.768033	5.533577	0.175716
1 5	2.715800	3.034631	3.013950	0.175716
2 5	0.477245	0.657958	0.820466	0.173158
3 5	1.540116	1.596497	1.415219	0.173147
4 5	2.683376	2.494611	2.496766	0.173147
5 5	-2.591738	-2.673920	-2.709411	0.173147
6 5	-2.885299	-2.049821	-1.968859	0.173147
7 5	-1.685347	-1.874237	-1.759651	0.173147
8 5	1.762811	1.969213	1.983631	0.173147
9 5	5.469728	5.481871	5.433663	0.173147
10 5	2.498080	2.215176	2.130636	0.173147
11 5	2.833613	2.720313	2.599917	0.173147
12 5	2.048597	2.009767	2.260383	0.173147
13 5	0.848944	1.551943	1.358802	0.173147
14 5	-0.099130	-0.118233	0.009404	0.173158
15 5	3.136872	3.621545	3.639475	0.175716
1 6	4.450724	4.528616	4.479051	0.175716
2 6	1.734696	1.512254	1.442451	0.173158
3 6	2.468628	2.470373	2.366727	0.173147
4 6	3.292462	3.676251	3.485033	0.173147
5 6	-3.793015	-3.855184	-3.580734	0.173147
6 6	-2.125749	-1.637413	-1.723214	0.173147
7 6	1.110032	0.858100	0.795642	0.173147
8 6	2.730643	2.710800	2.830417	0.173147
9 6	5.246077	6.090190	5.854375	0.173147
10 6	0.067565	-0.025937	0.155033	0.173147
11 6	1.409668	1.524446	1.454258	0.173147
12 6	2.870877	2.921055	2.815499	0.173147
13 6	0.509718	0.278465	0.394909	0.173147
14 6	0.914901	0.596927	0.492941	0.173158
15 6	3.659658	2.978502	2.800166	0.175716

(i,j)		X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1	7	4.663325	4.718497	4.639820	0.175716
2	7	0.866255	-0.434114	-0.194314	0.173158
3	7	1.408271	0.543806	0.610204	0.173147
4	7	2.431037	2.372699	2.242631	0.173147
5	7	-3.218578	-2.546162	-2.539237	0.173147
6	7	-2.053076	-2.391321	-2.124894	0.173147
7	7	2.136788	2.692481	2.607672	0.173147
8	7	4.202834	4.399781	4.193876	0.173147
9	7	4.527806	4.565724	4.373945	0.173147
10	7	-0.936523	-0.908308	-0.855764	0.173147
11	7	-1.691724	-0.792226	-0.626841	0.173147
12	7	1.939228	2.098638	1.975394	0.173147
13	7	-0.249195	-0.673773	-0.552258	0.173147
14	7	-0.180645	-1.319004	-1.359852	0.173158
15	7	-1.578310	-1.320057	-1.201563	0.175716
1	8	4.950517	5.454179	5.467435	0.175716
2	8	1.735909	1.810559	1.680239	0.173158
3	8	0.680859	0.646223	0.678492	0.173147
4	8	0.945482	0.445278	0.559823	0.173147
5	8	-0.412311	-0.791964	-0.790746	0.173147
6	8	0.180950	0.298041	0.244460	0.173147
7	8	2.842166	3.508742	3.187854	0.173147
8	8	1.602747	1.496752	1.697892	0.173147
9	8	-0.506876	-0.529515	-0.475970	0.173147
10	8	-2.480257	-2.101993	-1.950369	0.173147
11	8	-1.391746	-0.784978	-0.926305	0.173147
12	8	1.139510	0.829767	1.007753	0.173147
13	8	-0.246591	0.199374	0.032599	0.173147
14	8	-3.044809	-2.886991	-2.589677	0.173158
15	8	-3.260604	-3.597554	-3.400139	0.175716
1	9	6.517829	6.364485	6.223813	0.175716
2	9	1.643466	2.069445	2.105862	0.173158
3	9	1.097950	1.102451	1.100431	0.173147
4	9	0.297153	0.238127	0.281527	0.173147
5	9	1.252403	0.729741	0.693317	0.173147
6	9	1.382467	0.925288	0.614152	0.173147
7	9	0.682674	-0.187760	0.157154	0.173147
8	9	0.702638	1.036975	1.054689	0.173147
9	9	-2.911324	-3.360218	-3.280080	0.173147
10	9	-0.424517	-0.679706	-0.693967	0.173147
11	9	-1.461571	-1.437489	-1.276989	0.173147
12	9	2.766040	2.645324	2.428526	0.173147
13	9	0.919472	1.377511	1.327883	0.173147
14	9	1.161062	1.496240	1.363939	0.173158
15	9	-1.474076	-1.076483	-1.023549	0.175716
1	10	5.066681	5.279040	5.379075	0.175716
2	10	2.398480	2.455003	2.350488	0.173158
3	10	1.112467	1.646784	1.453580	0.173147
4	10	-0.733140	-0.129530	-0.002318	0.173147
5	10	0.923742	0.969117	0.832947	0.173147

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 10	-2.301466	-2.119271	-1.757141	0.173147
7 10	0.540056	1.038783	0.955411	0.173147
8 10	2.336503	3.037832	2.702800	0.173147
9 10	-4.059434	-4.560586	-4.290662	0.173147
10 10	1.747618	1.671799	1.446877	0.173147
11 10	-0.520071	-0.579491	-0.549998	0.173147
12 10	1.439323	1.053161	1.048650	0.173147
13 10	-0.047000	0.240555	0.257762	0.173147
14 10	3.153268	3.021048	2.928883	0.173158
15 10	1.801906	1.248829	1.004144	0.175716
1 11	5.200778	5.572349	5.464753	0.175716
2 11	0.700298	0.730090	0.783691	0.173158
3 11	-1.752462	-1.045512	-0.849491	0.173147
4 11	-0.113598	0.159780	0.212409	0.173147
5 11	1.073313	1.055686	0.940259	0.173147
6 11	-0.810886	-0.492874	-0.413373	0.173147
7 11	0.774427	0.745980	0.814803	0.173147
8 11	0.667794	0.401388	0.600608	0.173147
9 11	-4.256761	-4.757324	-4.702361	0.173147
10 11	1.188914	1.040265	1.107832	0.173147
11 11	-0.446754	-0.623123	-0.645821	0.173147
12 11	-0.335805	-1.045615	-0.951531	0.173147
13 11	-1.896477	-2.125102	-1.865239	0.173147
14 11	2.228810	1.847806	1.702144	0.173158
15 11	-0.861164	-1.617185	-1.398470	0.175716
1 12	5.354785	5.210977	5.076009	0.175716
2 12	0.525775	0.672325	0.776251	0.173158
3 12	0.115824	0.039552	0.196525	0.173147
4 12	1.771480	2.441880	2.204293	0.173147
5 12	1.081651	0.936434	1.040533	0.173147
6 12	2.498181	3.006305	2.826046	0.173147
7 12	3.261475	3.802892	3.699306	0.173147
8 12	2.201032	3.144071	2.904099	0.173147
9 12	-2.733236	-2.374834	-2.470867	0.173147
10 12	2.954685	3.388874	3.228217	0.173147
11 12	0.236108	0.893256	0.879749	0.173147
12 12	0.384083	0.083742	0.196096	0.173147
13 12	0.204221	0.546608	0.361822	0.173147
14 12	1.777615	1.597079	1.598518	0.173158
15 12	-0.693838	-0.623532	-0.543026	0.175716
1 13	4.044871	3.635299	3.693018	0.175717
2 13	2.860570	2.053560	2.018352	0.173159
3 13	2.478425	2.684627	2.561723	0.173147
4 13	2.128292	1.640840	1.641819	0.173147
5 13	-0.226666	-0.209300	0.075679	0.173147

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 13	2.278092	2.397945	2.157062	0.173147
7 13	2.166856	2.026551	2.158851	0.173147
8 13	-0.902468	-0.661992	-0.564902	0.173147
9 13	-5.633426	-5.203928	-4.952505	0.173147
10 13	1.880781	1.686412	1.606602	0.173147
11 13	0.399469	1.060540	0.934524	0.173147
12 13	0.770136	1.370116	1.336095	0.173147
13 13	0.594662	0.359955	0.293117	0.173147
14 13	0.707030	0.964431	1.130464	0.173159
15 13	1.037011	1.268357	1.171869	0.175717
1 14	2.709518	3.038597	3.040798	0.175757
2 14	2.163806	2.557711	2.644371	0.173220
3 14	3.933929	3.837133	3.624959	0.173208
4 14	1.047408	1.839649	1.807417	0.173208
5 14	1.880644	2.327353	2.345250	0.173208
6 14	0.338981	0.619696	0.770894	0.173208
7 14	3.702429	3.171865	3.154970	0.173208
8 14	-0.950579	-0.661404	-0.745015	0.173208
9 14	-4.027186	-3.783536	-3.788032	0.173208
10 14	-0.210480	-0.297191	-0.154588	0.173208
11 14	-0.393181	-1.119339	-0.965028	0.173208
12 14	1.156109	1.422708	1.215719	0.173208
13 14	-0.853163	-0.643251	-0.510388	0.173208
14 14	1.771989	2.209265	2.157937	0.173220
15 14	2.751313	2.429512	2.232267	0.175757
1 15	2.849064	2.613831	2.640832	0.183585
2 15	4.349098	3.664304	3.486165	0.181651
3 15	2.273647	1.951333	2.068983	0.181646
4 15	1.311287	0.743600	0.903642	0.181646
5 15	4.966369	5.642505	5.270374	0.181646
6 15	1.530654	1.339646	1.461264	0.181646
7 15	4.542632	4.895004	4.640556	0.181646
8 15	-1.694414	-1.868427	-1.639006	0.181646
9 15	-1.921006	-2.174508	-2.257422	0.181646
10 15	0.506533	0.770858	0.704469	0.181646
11 15	-0.653809	-0.722584	-0.738990	0.181646
12 15	-1.218652	-0.422465	-0.301761	0.181646
13 15	-1.411769	-0.859022	-0.851151	0.181646
14 15	1.314279	1.865621	1.775317	0.181651
15 15	0.047197	0.006456	0.160770	0.183585

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.35$, $\beta = 0.75$ y $\sigma_\eta^2 = 2$, datos Y_{ij} de la ecuación de observación con todos los $m_{ij} = 1$ y $\sigma_\epsilon^2 = 0.2$, datos de X alisados y errores esperados con un 75% de los datos observados.

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1 1	3.565483	4.062298	3.801656	0.184411
2 1	1.341963	1.918847	1.970841	0.182445
3 1	2.322185	2.305162	2.159412	0.182444
4 1	-1.347617	-1.227081	-1.006249	0.184127
5 1	-2.701298		-1.460068	2.079375
6 1	-3.519091	-3.701221	-3.569372	0.184157
7 1	-5.418998	-4.918092	-4.769477	0.182455
8 1	-2.056990	-1.903632	-1.659719	0.182441
9 1	2.376616	2.316155	2.146751	0.182443
10 1	0.330162	-0.563863	-0.561603	0.182907
11 1	-4.606561	-3.751174	-3.577731	0.188697
12 1	0.978110		0.940067	2.136987
13 1	-0.118302	0.271384	0.131699	0.185930
14 1	-1.212180		-0.401398	2.079032
15 1	0.500941	0.531147	0.616971	0.186151
1 2	1.106009	0.535414	0.730923	0.177320
2 2	1.582366	1.324249	1.372721	0.174750
3 2	2.384358	1.923487	1.781174	0.174763
4 2	1.457827	2.011415	1.999734	0.175364
5 2	0.389788	0.890241	0.770039	0.183055
6 2	-0.585789	-0.261146	-0.360343	0.181891
7 2	-0.380235	-1.195831	-1.261054	0.175252
8 2	2.650302	2.892981	2.596079	0.174763
9 2	2.328302	2.638155	2.738412	0.174740
10 2	0.870360	0.063864	0.150047	0.177017
11 2	-2.310778		-2.166400	1.425486
12 2	1.946970	1.606965	1.567683	0.183555
13 2	-1.160373	-1.816912	-1.532971	0.176053
14 2	-0.625260	-0.721370	-0.719542	0.182443
15 2	2.424585	2.023899	2.005998	0.177985
1 3	0.254407	0.928657	0.989126	0.181900
2 3	1.462085	2.024453	1.864812	0.175708
3 3	-0.886270	-1.007741	-0.670764	0.181055
4 3	1.974845	3.280179	2.957314	0.175215
5 3	-0.739446	-0.682381	-0.606327	0.176948
6 3	-0.953320		-1.123433	1.763223
7 3	-2.035827	-1.877644	-1.869095	0.177432
8 3	1.961053	1.892569	1.915356	0.181056
9 3	3.484118	3.474646	3.511648	0.175214
10 3	2.604234	2.438938	2.307230	0.175613

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
11 3	0.230612	0.084809	0.110274	0.186557
12 3	3.303419	3.950325	3.872642	0.175691
13 3	1.035067	1.579562	1.440764	0.174704
14 3	0.399183	0.731504	0.775636	0.175163
15 3	4.460101	4.205260	4.067649	0.181739
1 4	1.721876		1.887445	1.602539
2 4	1.742069	1.735557	1.709026	0.183395
3 4	-1.712308		-0.313202	1.751707
4 4	1.135263	0.880484	1.071760	0.177371
5 4	-1.765706	-2.122092	-2.149376	0.181033
6 4	-0.858743		-1.942541	1.791188
7 4	-1.597033	-2.488470	-2.409613	0.179310
8 4	1.422393		1.686926	1.751161
9 4	4.818954	4.404212	4.372800	0.181416
10 4	2.257835	2.471390	2.456892	0.177348
11 4	2.513007		1.045876	1.753318
12 4	3.991704	4.082114	3.942101	0.176894
13 4	2.124664	1.293349	1.375751	0.174691
14 4	0.856867	0.925426	0.860074	0.177259
15 4	5.141075		3.844748	1.559730
1 5	2.715800	3.034631	2.976066	0.189377
2 5	0.477245		1.776182	1.628228
3 5	1.540116		1.163908	1.922403
4 5	2.683376	2.494611	2.458978	0.179623
5 5	-2.591738		-2.790942	1.413456
6 5	-2.885299	-2.049821	-2.052145	0.182396
7 5	-1.685347	-1.874237	-1.748940	0.177797
8 5	1.762811		1.900635	1.907127
9 5	5.469728		5.055519	1.533286
10 5	2.498080	2.215176	2.087032	0.179516
11 5	2.833613		1.066724	1.751567
12 5	2.048597	2.009767	2.209864	0.176893
13 5	0.848944	1.551943	1.367724	0.174692
14 5	-0.099130	-0.118233	0.033856	0.175147
15 5	3.136872	3.621545	3.539500	0.181749
1 6	4.450724		3.246128	2.259309
2 6	1.734696	1.512254	1.485311	0.181975
3 6	2.468628	2.470373	2.362845	0.186322
4 6	3.292462	3.676251	3.488295	0.176048
5 6	-3.793015	-3.855184	-3.618503	0.180322
6 6	-2.125749	-1.637413	-1.637228	0.182292
7 6	1.110032	0.858100	0.784145	0.175846
8 6	2.730643	2.710800	2.844080	0.181483
9 6	5.246077	6.090190	5.769159	0.184799
10 6	0.067565	-0.025937	0.207560	0.176026
11 6	1.409668	1.524446	1.355190	0.181088
12 6	2.870877	2.921055	2.854501	0.175203
13 6	0.509718	0.278465	0.382937	0.175179
14 6	0.914901	0.596927	0.501310	0.179629
15 6	3.659658	2.978502	2.854905	0.181750

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1 7	4.663325		3.157491	2.790703
2 7	0.866255	-0.434114	-0.216443	0.179594
3 7	1.408271		0.979605	1.384228
4 7	2.431037	2.372699	2.272468	0.176920
5 7	-3.218578	-2.546162	-2.468462	0.176948
6 7	-2.053076		-0.797632	1.919883
7 7	2.136788	2.692481	2.695447	0.176931
8 7	4.202834	4.399781	4.151740	0.176935
9 7	4.527806		3.244621	1.383857
10 7	-0.936523	-0.908308	-0.891271	0.177289
11 7	-1.691724	-0.792226	-0.703830	0.179537
12 7	1.939228	2.098638	2.019542	0.175462
13 7	-0.249195	-0.673773	-0.621448	0.181264
14 7	-0.180645		-1.156570	1.589721
15 7	-1.578310		-0.293335	1.764645
1 8	4.950517		4.814543	2.185200
2 8	1.735909	1.810559	1.679778	0.178119
3 8	0.680859	0.646223	0.709666	0.184783
4 8	0.945482	0.445278	0.569243	0.176056
5 8	-0.412311	-0.791964	-0.897866	0.182747
6 8	0.180950		-0.343470	2.433622
7 8	2.842166	3.508742	3.296369	0.181434
8 8	1.602747	1.496752	1.644983	0.176056
9 8	-0.506876	-0.529515	-0.453132	0.186375
10 8	-2.480257	-2.101993	-2.016834	0.177872
11 8	-1.391746		-1.852003	1.380795
12 8	1.139510	0.829767	0.904811	0.179261
13 8	-0.246591		-1.093772	1.408129
14 8	-3.044809	-2.886991	-2.671453	0.181045
15 8	-3.260604	-3.597554	-3.367876	0.181742
1 9	6.517829	6.364485	6.202243	0.184281
2 9	1.643466	2.069445	2.080457	0.181813
3 9	1.097950		1.188892	1.405206
4 9	0.297153	0.238127	0.245027	0.179626
5 9	1.252403		-0.733684	1.943520
6 9	1.382467		-1.254268	2.271229
7 9	0.682674		1.696323	1.556066
8 9	0.702638	1.036975	1.156751	0.179640
9 9	-2.911324		-1.656117	1.766140
10 9	-0.424517	-0.679706	-0.709505	0.181990
11 9	-1.461571	-1.437489	-1.323666	0.180041
12 9	2.766040	2.645324	2.480638	0.175438
13 9	0.919472	1.377511	1.277641	0.180053
14 9	1.161062	1.496240	1.337444	0.180113
15 9	-1.474076	-1.076483	-0.997884	0.177727
1 10	5.066681	5.279040	5.337941	0.179658
2 10	2.398480		1.705919	1.403913
3 10	1.112467	1.646784	1.439520	0.181027
4 10	-0.733140	-0.129530	-0.043725	0.177464
5 10	0.923742		-0.506997	1.811692

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 10	-2.301466	-2.119271	-1.932362	0.187894
7 10	0.540056	1.038783	1.074758	0.180618
8 10	2.336503	3.037832	2.812663	0.181748
9 10	-4.059434		-2.285487	1.960938
10 10	1.747618		0.518380	1.566308
11 10	-0.520071	-0.579491	-0.608982	0.181378
12 10	1.439323	1.053161	1.056267	0.175160
13 10	-0.047000	0.240555	0.214458	0.176987
14 10	3.153268		2.149288	1.407982
15 10	1.801906	1.248829	1.085440	0.183304
1 11	5.200778	5.572349	5.480388	0.177685
2 11	0.700298	0.730090	0.736681	0.179654
3 11	-1.752462	-1.045512	-0.851144	0.175131
4 11	-0.113598	0.159780	0.234295	0.175256
5 11	1.073313	1.055686	0.893911	0.182450
6 11	-0.810886		0.123851	1.418328
7 11	0.774427	0.745980	0.887898	0.179370
8 11	0.667794		1.503824	1.417317
9 11	-4.256761	-4.757324	-4.508747	0.182362
10 11	1.188914	1.040265	0.990194	0.181203
11 11	-0.446754		-0.818757	1.436080
12 11	-0.335805	-1.045615	-0.914884	0.181186
13 11	-1.896477	-2.125102	-1.876834	0.175480
14 11	2.228810	1.847806	1.728713	0.181268
15 11	-0.861164		0.283726	1.590359
1 12	5.354785	5.210977	5.107004	0.181740
2 12	0.525775	0.672325	0.747012	0.179544
3 12	0.115824	0.039552	0.139898	0.179503
4 12	1.771480	2.441880	2.211239	0.175568
5 12	1.081651	0.936434	1.135337	0.179995
6 12	2.498181	3.006305	2.783362	0.185004
7 12	3.261475	3.802892	3.698408	0.175967
8 12	2.201032	3.144071	2.978531	0.179689
9 12	-2.733236	-2.374834	-2.457762	0.175127
10 12	2.954685	3.388874	3.240528	0.175735
11 12	0.236108	0.893256	0.876647	0.187190
12 12	0.384083		0.964009	1.440594
13 12	0.204221	0.546608	0.410842	0.177039
14 12	1.777615	1.597079	1.564715	0.175191
15 12	-0.693838	-0.623532	-0.448211	0.181857
1 13	4.044871		3.919949	1.790946
2 13	2.860570		1.516962	1.760462
3 13	2.478425		1.696530	1.559960
4 13	2.128292	1.640840	1.663069	0.179908
5 13	-0.226666		1.450251	1.535946

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 13	2.278092		1.207635	1.568460
7 13	2.166856	2.026551	2.087604	0.181391
8 13	-0.902468	-0.661992	-0.567644	0.175126
9 13	-5.633426	-5.203928	-4.971120	0.174739
10 13	1.880781	1.686412	1.576622	0.177020
11 13	0.399469		0.391086	1.798171
12 13	0.770136	1.370116	1.279993	0.186128
13 13	0.594662	0.359955	0.305130	0.175635
14 13	0.707030	0.964431	1.115457	0.174708
15 13	1.037011	1.268357	1.183866	0.177288
1 14	2.709518	3.038597	3.097478	0.182504
2 14	2.163806	2.557711	2.505817	0.187791
3 14	3.933929	3.837133	3.620131	0.180251
4 14	1.047408	1.839649	1.803708	0.175616
5 14	1.880644	2.327353	2.432729	0.180295
6 14	0.338981	0.619696	0.739911	0.188385
7 14	3.702429		2.974070	1.454583
8 14	-0.950579	-0.661404	-0.733689	0.177753
9 14	-4.027186	-3.783536	-3.832789	0.182438
10 14	-0.210480	-0.297191	-0.128779	0.177918
11 14	-0.393181		-0.692320	1.929789
12 14	1.156109		0.079399	1.586454
13 14	-0.853163	-0.643251	-0.544936	0.183712
14 14	1.771989	2.209265	2.150796	0.175421
15 14	2.751313	2.429512	2.245770	0.177305
1 15	2.849064	2.613831	2.584211	0.186164
2 15	4.349098		1.721394	2.083060
3 15	2.273647	1.951333	2.003939	0.184150
4 15	1.311287	0.743600	0.896243	0.182464
5 15	4.966369	5.642505	5.331001	0.184220
6 15	1.530654		2.642662	2.145551
7 15	4.542632	4.895004	4.672087	0.188803
8 15	-1.694414	-1.868427	-1.663661	0.184627
9 15	-1.921006		-2.993593	2.078781
10 15	0.506533	0.770858	0.677770	0.184848
11 15	-0.653809	-0.722584	-0.691963	0.189602
12 15	-1.218652	-0.422465	-0.385228	0.188873
13 15	-1.411769		-0.503934	2.145651
14 15	1.314279	1.865621	1.786557	0.184209
15 15	0.047197	0.006456	0.152183	0.184414

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.35$, $\beta = 0.75$ y $\sigma_\eta^2 = 2$, datos Y_{ij} de la ecuación de observación con todos los $m_{ij} = 1$ y $\sigma_\epsilon^2 = 0.2$, datos de X alisados y errores esperados con un 50% de los datos observados.

(i,j)		X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1	1	3.565483	4.062298	3.805735	0.185132
2	1	1.341963	1.918847	1.918189	0.189783
3	1	2.322185		0.811252	2.055992
4	1	-1.347617	-1.227081	-1.127480	0.183815
5	1	-2.701298	-3.426069	-3.208048	0.184089
6	1	-3.519091		-2.324967	3.308813
7	1	-5.418998		-2.208763	4.073344
8	1	-2.056990	-1.903632	-1.745826	0.188904
9	1	2.376616	2.316155	2.164622	0.186869
10	1	0.330162	-0.563863	-0.566008	0.182747
11	1	-4.606561	-3.751174	-3.585409	0.188158
12	1	0.978110		0.276319	2.111964
13	1	-0.118302	0.271384	0.212465	0.188607
14	1	-1.212180	-1.721247	-1.622739	0.189956
15	1	0.500941	0.531147	0.513313	0.191253
1	2	1.106009	0.535414	0.761324	0.185558
2	2	1.582366		1.167635	1.796080
3	2	2.384358	1.923487	1.891947	0.190104
4	2	1.457827	2.011415	1.914558	0.179845
5	2	0.389788	0.890241	0.652191	0.177850
6	2	-0.585789		-0.345998	2.298539
7	2	-0.380235		-0.896426	4.018506
8	2	2.650302		0.130255	1.813236
9	2	2.328302		2.513394	1.552671
10	2	0.870360	0.063864	0.143506	0.180433
11	2	-2.310778		-2.071580	1.374091
12	2	1.946970	1.606965	1.627406	0.185025
13	2	-1.160373		0.687197	1.629984
14	2	-0.625260		-0.708018	2.560396
15	2	2.424585		1.165216	2.622692
1	3	0.254407		1.439660	2.355782
2	3	1.462085		1.359982	2.272537
3	3	-0.886270		1.608632	2.338298
4	3	1.974845		1.329945	1.554653
5	3	-0.739446	-0.682381	-0.674206	0.179378
6	3	-0.953320	-0.914847	-0.899303	0.190935
7	3	-2.035827		-0.748490	2.944357
8	3	1.961053	1.892569	1.779266	0.185924
9	3	3.484118	3.474646	3.569489	0.179147
10	3	2.604234	2.438938	2.314126	0.175509

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
11 3	0.230612	0.084809	0.106323	0.186463
12 3	3.303419	3.950325	3.838752	0.175769
13 3	1.035067	1.579562	1.573386	0.181897
14 3	0.399183		0.312877	3.147013
15 3	4.460101		2.007280	3.251846
1 4	1.721876		2.253785	2.102012
2 4	1.742069	1.735557	1.685287	0.189036
3 4	-1.712305		1.490066	2.294454
4 4	1.135263	0.880484	0.903145	0.186259
5 4	-1.765706	-2.122092	-1.988070	0.184598
6 4	-0.858743		-1.622996	2.139469
7 4	-1.597033		-1.421806	2.451695
8 4	1.422393		1.773658	1.546944
9 4	4.818954	4.404212	4.425478	0.177782
10 4	2.257835	2.471390	2.402657	0.180998
11 4	2.513007		0.672804	1.807841
12 4	3.991704	4.082114	3.961087	0.177217
13 4	2.124664	1.293349	1.319188	0.181481
14 4	0.856867		0.733138	3.274968
15 4	5.141075		2.724396	2.572851
1 5	2.715800	3.034631	3.003161	0.190388
2 5	0.477245		1.612229	1.713029
3 5	1.540116		1.853401	2.368022
4 5	2.683376		1.986272	1.884253
5 5	-2.591738		-0.880758	2.569507
6 5	-2.885299		-1.262087	3.010637
7 5	-1.685347	-1.874237	-1.707641	0.184292
8 5	1.762811	1.969213	2.006567	0.184128
9 5	5.469728	5.481871	5.361958	0.178199
10 5	2.498080		1.191648	1.501713
11 5	2.833613		-0.205934	2.461919
12 5	2.048597	2.009767	2.151164	0.180709
13 5	0.848944		0.432373	1.515752
14 5	-0.099130		0.851340	2.893567
15 5	3.136872	3.621545	3.509564	0.186283
1 6	4.450724		3.474082	2.101879
2 6	1.734696	1.512254	1.480358	0.181493
3 6	2.468628	2.470373	2.417264	0.181879
4 6	3.292462	3.676251	3.535948	0.188637
5 6	-3.793015		0.836412	3.150318
6 6	-2.125749		0.445494	3.015958
7 6	1.110032	0.858100	0.860179	0.180191
8 6	2.730643		3.696346	1.317742
9 6	5.246077	6.090190	5.907267	0.177119
10 6	0.067565	-0.025937	0.099820	0.180995
11 6	1.409668		-0.196370	1.807575
12 6	2.870877	2.921055	2.823175	0.177211
13 6	0.509718	0.278465	0.318894	0.181082
14 6	0.914901		0.619098	1.916422
15 6	3.659658	2.978502	2.797765	0.180392

(i,j)		X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1	7	4.663325		3.406481	2.088648
2	7	0.866255	-0.434114	-0.262978	0.181499
3	7	1.408271	0.543806	0.571823	0.177877
4	7	2.431037		2.055088	2.144973
5	7	-3.218578		0.883846	3.788599
6	7	-2.053076		1.128152	2.304761
7	7	2.136788	2.692481	2.555834	0.181737
8	7	4.202834	4.399781	4.322350	0.179638
9	7	4.527806	4.565724	4.396168	0.179005
10	7	-0.936523	-0.908308	-0.791836	0.175344
11	7	-1.691724	-0.792226	-0.809948	0.186150
12	7	1.939228	2.098638	2.030375	0.175822
13	7	-0.249195	-0.673773	-0.652176	0.182366
14	7	-0.180645	-1.319004	-1.262573	0.187380
15	7	-1.578310	-1.320057	-1.089269	0.184094
1	8	4.950517	5.454179	5.308401	0.190729
2	8	1.735909		1.354015	1.495185
3	8	0.680859	0.646223	0.671096	0.185200
4	8	0.945482		1.346222	2.714264
5	8	-0.412311		0.598194	3.533016
6	8	0.180950	0.298041	0.282290	0.190829
7	8	2.842166		0.769569	1.434651
8	8	1.602747	1.496752	1.575630	0.179943
9	8	-0.506876		-0.264187	1.317533
10	8	-2.480257	-2.101993	-2.006736	0.179549
11	8	-1.391746		-1.994339	1.320285
12	8	1.139510	0.829767	0.935832	0.180252
13	8	-0.246591		-0.418914	2.192085
14	8	-3.044809		-0.206456	1.829019
15	8	-3.260604		-0.266283	2.348967
1	9	6.517829		5.059223	2.257052
2	9	1.643466	2.069445	1.986426	0.189076
3	9	1.097950		0.662865	2.153364
4	9	0.297153		0.712600	2.462501
5	9	1.252403		0.415187	3.541473
6	9	1.382467		-0.368570	2.070798
7	9	0.682674	-0.187760	-0.104654	0.187208
8	9	0.702638	1.036975	0.933919	0.181248
9	9	-2.911324	-3.360218	-3.138459	0.183937
10	9	-0.424517	-0.679706	-0.784694	0.179719
11	9	-1.461571	-1.437489	-1.330819	0.185425
12	9	2.766040	2.645324	2.444322	0.178180
13	9	0.919472		1.009675	2.845521
14	9	1.161062	1.496240	1.435263	0.189657
15	9	-1.474076		0.828562	2.112660
1	10	5.066681		4.522863	3.221669
2	10	2.398480		1.271016	2.378267
3	10	1.112467		0.026721	2.739572
4	10	-0.733140	-0.129530	-0.069023	0.190756
5	10	0.923742		0.419262	2.400909

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 10	-2.301466		-0.462522	2.295768
7 10	0.540056		0.357051	2.313386
8 10	2.336503		0.757082	2.311425
9 10	-4.059434		-3.664736	1.885963
10 10	1.747618		0.156395	1.825230
11 10	-0.520071		-0.876987	1.904459
12 10	1.439323	1.053161	1.036884	0.180861
13 10	-0.047000		0.563301	3.645400
14 10	3.153268		1.368698	2.321395
15 10	1.801906	1.248829	1.097794	0.185238
1 11	5.200778		4.533916	2.623864
2 11	0.700298		0.945682	2.414641
3 11	-1.752462		-0.047867	2.476972
4 11	-0.113598		0.284093	2.091693
5 11	1.073313	1.055686	1.017434	0.189508
6 11	-0.810886	-0.492874	-0.333151	0.183882
7 11	0.774427		0.967699	1.933331
8 11	0.667794		0.678224	1.908740
9 11	-4.256761	-4.757324	-4.569178	0.181484
10 11	1.188914	1.040265	0.973147	0.181004
11 11	-0.446754		-0.810715	1.659853
12 11	-0.335805	-1.045615	-0.915299	0.180277
13 11	-1.896477		-0.665594	4.008847
14 11	2.228810		-0.044020	3.067249
15 11	-0.861164	-1.617185	-1.422010	0.184734
1 12	5.354785	5.210977	5.013609	0.183648
2 12	0.525775	0.672325	0.817882	0.181232
3 12	0.115824	0.039552	0.099998	0.189220
4 12	1.771480		1.126079	2.021813
5 12	1.081651		2.653275	1.885061
6 12	2.498181	3.006305	2.814590	0.182600
7 12	3.261475	3.802892	3.758028	0.180939
8 12	2.201032	3.144071	2.905166	0.180598
9 12	-2.733236	-2.374834	-2.303547	0.179379
10 12	2.954685	3.388874	3.201403	0.174736
11 12	0.236108	0.893256	0.907896	0.181257
12 12	0.384083	0.083742	0.095496	0.184893
13 12	0.204221		0.210339	4.196686
14 12	1.777615		0.973465	3.420534
15 12	-0.693838		-0.076769	1.891553
1 13	4.044871	3.635299	3.698419	0.184196
2 13	2.860570	2.053560	1.950043	0.183495
3 13	2.478425		1.244633	1.787318
4 13	2.128292	1.640840	1.614128	0.184844
5 13	-0.226666		2.387773	1.886492

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 13	2.278092		1.250883	1.556981
7 13	2.166856	2.026551	2.078198	0.181106
8 13	-0.902468	-0.661992	-0.501394	0.177097
9 13	-5.633426		-3.363863	1.315330
10 13	1.880781	1.686412	1.647446	0.177609
11 13	0.399469	1.060540	1.005734	0.184137
12 13	0.770136		0.448946	2.296233
13 13	0.594662		0.779114	4.481049
14 13	0.707030		1.919956	2.353563
15 13	1.037011	1.268357	1.183096	0.188475
1 14	2.709518		2.927914	2.827757
2 14	2.163806		1.832999	2.615926
3 14	3.933929		1.716879	1.936089
4 14	1.047408	1.839649	1.687529	0.178080
5 14	1.880644	2.327353	2.486603	0.181222
6 14	0.338981	0.619696	0.733565	0.187922
7 14	3.702429		2.801034	1.403210
8 14	-0.950579	-0.661404	-0.798926	0.177302
9 14	-4.027186	-3.783536	-3.665601	0.179663
10 14	-0.210480	-0.297191	-0.213318	0.184735
11 14	-0.393181		0.113932	2.314053
12 14	1.156109		0.169449	3.771739
13 14	-0.853163		0.794911	4.360064
14 14	1.771989	2.209265	2.136185	0.192061
15 14	2.751313		0.883149	1.747147
1 15	2.849064		2.313200	4.146785
2 15	4.349098		1.655489	3.506195
3 15	2.273647	1.951333	1.880836	0.190379
4 15	1.311287	0.743600	0.965846	0.182379
5 15	4.966369	5.642505	5.332550	0.183789
6 15	1.530654		2.943968	2.027727
7 15	4.542632	4.895004	4.657120	0.188237
8 15	-1.694414	-1.868427	-1.609682	0.182032
9 15	-1.921006	-2.174508	-2.264277	0.184174
10 15	0.506533		0.048706	2.325350
11 15	-0.653809		0.158813	3.891081
12 15	-1.218652		0.107383	4.745389
13 15	-1.411769		0.468154	5.010052
14 15	1.314279		1.294884	2.476350
15 15	0.047197	0.006456	0.050070	0.190782

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.35$, $\beta = 0.75$ y $\sigma_\eta^2 = 2$, datos Y_{ij} de la ecuación de observación con todos los $m_{ij} = 1$ y $\sigma_\epsilon^2 = 0.2$, datos de X alisados y errores esperados con un 25% de los datos observados.

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1 1	3.565483	4.062298	3.970665	0.192842
2 1	1.341963		2.119376	3.471260
3 1	2.322185	2.305162	2.255390	0.191750
4 1	-1.347617		1.171041	2.843779
5 1	-2.701298	-3.426069	-3.337999	0.191550
6 1	-3.519091		-3.482914	2.838799
7 1	-5.418998	-4.918092	-4.612717	0.185107
8 1	-2.056990		-0.483889	3.289393
9 1	2.376616	2.316155	2.209566	0.185440
10 1	0.330162		1.663318	3.892690
11 1	-4.606561		1.301211	4.841953
12 1	0.978110		1.070336	4.323420
13 1	-0.118302		-1.296034	2.652100
14 1	-1.212180		-0.076580	4.574313
15 1	0.500941		1.520744	2.658644
1 2	1.106009		3.615894	2.578574
2 2	1.582366		1.916701	3.952941
3 2	2.384358		2.432902	2.299279
4 2	1.457827		2.241416	2.355039
5 2	0.389788		-1.987673	2.155585
6 2	-0.585789		-1.589508	2.208694
7 2	-0.380235	-1.195831	-1.333311	0.184213
8 2	2.650302		1.194501	2.891972
9 2	2.328302	2.638155	2.613472	0.184716
10 2	0.870360		2.006405	3.338688
11 2	-2.310778		1.593609	4.283601
12 2	1.946970		1.325970	3.197695
13 2	-1.160373	-1.816912	-1.656149	0.190239
14 2	-0.625260		-0.105360	3.599973
15 2	2.424585	2.023899	1.920443	0.192674
1 3	0.254407		3.409131	3.712336
2 3	1.462085		1.721315	3.597873
3 3	-0.886270		2.561584	2.570964
4 3	1.974845	3.280179	3.094984	0.184889
5 3	-0.739446		-1.381508	2.199620
6 3	-0.953320	-0.914847	-0.957432	0.188986
7 3	-2.035827		-0.299813	2.167862
8 3	1.961053		1.687227	2.345107
9 3	3.484118		2.339324	2.350612
10 3	2.604234		2.240217	2.483916

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
11 3	0.230612		2.269649	3.572959
12 3	3.303419		2.581736	2.461837
13 3	1.035067		-0.188270	1.721886
14 3	0.399183		0.289428	4.141813
15 3	4.460101		1.534086	2.501431
1 4	1.721876		2.914946	4.023339
2 4	1.742069		0.737336	2.483731
3 4	-1.712305		1.194960	1.993533
4 4	1.135263	0.880484	1.087433	0.179166
5 4	-1.765706		-2.740864	1.903402
6 4	-0.858743		-1.861691	1.618491
7 4	-1.597033		-0.083811	2.914048
8 4	1.422393	1.842382	1.841728	0.189418
9 4	4.818954		1.962261	2.817880
10 4	2.257835	2.471390	2.472993	0.184982
11 4	2.513007		3.004298	1.951445
12 4	3.991704	4.082114	3.944854	0.190520
13 4	2.124664	1.293349	1.249244	0.188341
14 4	0.856867		0.689653	3.708890
15 4	5.141075		1.227576	3.389238
1 5	2.715800		3.080928	3.609220
2 5	0.477245	0.657958	0.715713	0.190681
3 5	1.540116	1.596497	1.600963	0.188777
4 5	2.683376	2.494611	2.299696	0.182385
5 5	-2.591738	-2.673920	-2.571079	0.187605
6 5	-2.885299	-2.049821	-2.005545	0.181273
7 5	-1.685347		0.512715	3.057540
8 5	1.762811		2.156062	1.619296
9 5	5.469728		1.459089	3.297168
10 5	2.498080	2.215176	2.223161	0.183698
11 5	2.833613	2.720313	2.625503	0.187978
12 5	2.048597		3.084518	2.268058
13 5	0.848944		0.957238	2.406333
14 5	-0.099130		0.156776	3.588710
15 5	3.136872		0.464929	3.475214
1 6	4.450724		3.609371	3.565847
2 6	1.734696		1.086774	2.394994
3 6	2.468628		1.586325	2.483191
4 6	3.292462		1.977799	2.433390
5 6	-3.793015		-2.085659	1.799089
6 6	-2.125749	-1.637413	-1.571839	0.182997
7 6	1.110032		1.436199	2.617609
8 6	2.730643	2.710800	2.614810	0.189271
9 6	5.246077		0.796573	3.350546
10 6	0.067565		1.497430	2.306026
11 6	1.409668		1.038134	1.714077
12 6	2.870877		1.664593	3.465114
13 6	0.509718		0.323509	3.238660
14 6	0.914901		-0.580733	2.545300
15 6	3.659658		-0.386991	2.520281

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
1 7	4.663325		4.365792	2.543277
2 7	0.866255		1.567564	2.977871
3 7	1.408271		1.751098	3.398594
4 7	2.431037		1.934564	3.470698
5 7	-3.218578		-1.405082	2.117794
6 7	-2.053076		-0.648869	1.719421
7 7	2.136788		2.187694	2.314945
8 7	4.202834		2.039384	2.348359
9 7	4.527806		-0.458750	3.356927
10 7	-0.936523		0.511161	2.820401
11 7	-1.691724	-0.792226	-0.677112	0.190091
12 7	1.939228		0.237243	3.415121
13 7	-0.249195		-0.346348	2.804261
14 7	-0.180645	-1.319004	-1.379922	0.184023
15 7	-1.578310	-1.320057	-1.276316	0.190884
1 8	4.950517	5.454179	5.369461	0.191031
2 8	1.735909		2.142409	2.126233
3 8	0.680859		2.025749	3.194798
4 8	0.945482		2.021778	3.413043
5 8	-0.412311	-0.791964	-0.761355	0.189821
6 8	0.180950	0.298041	0.314888	0.186458
7 8	2.842166	3.508742	3.203921	0.183705
8 8	1.602747		1.841893	2.808938
9 8	-0.506876		-1.250627	2.414197
10 8	-2.480257		0.459087	3.132204
11 8	-1.891746		-0.753009	2.492516
12 8	1.139510		-0.389487	3.680643
13 8	-0.246591		-0.829402	2.051739
14 8	-3.044809	-2.886991	-2.550469	0.179184
15 8	-3.260604		-1.920099	2.069037
1 9	6.517829		5.309724	1.927252
2 9	1.643466	2.069445	2.018779	0.188628
3 9	1.097950		1.819925	2.149103
4 9	0.297153		1.620222	3.632304
5 9	1.252403		-1.005585	2.466357
6 9	1.382467		-0.740837	2.061993
7 9	0.682674	-0.187760	0.032796	0.178872
8 9	0.702638		-0.519780	2.586278
9 9	-2.911324	-3.360218	-3.168445	0.191650
10 9	-0.424517		0.173541	2.184027
11 9	-1.461571		-0.434541	3.431103
12 9	2.766040		0.264942	3.118995
13 9	0.919472	1.377511	1.321194	0.190744
14 9	1.161062	1.496240	1.243752	0.183938
15 9	-1.474076		0.086788	2.877402
1 10	5.066681		5.229774	2.113779
2 10	2.398480		1.421516	1.566010
3 10	1.112467	1.646784	1.650343	0.190354
4 10	-0.733140		1.739126	3.065023
5 10	0.923742		-0.212065	3.555066

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.
6 10	-2.301466		0.348474	2.859444
7 10	0.540056	1.038783	0.999148	0.184861
8 10	2.336503		0.352845	2.099772
9 10	-4.059434		-1.593414	2.173759
10 10	1.747618	1.671799	1.549390	0.184680
11 10	-0.520071		0.140588	3.555910
12 10	1.439323		-0.045628	2.354239
13 10	-0.047000		0.921112	2.471595
14 10	3.153268		0.944180	2.492721
15 10	1.801906		-0.149619	3.274853
1 11	5.200778	5.572349	5.435270	0.182834
2 11	0.700298	0.730090	0.891580	0.186858
3 11	-1.752462		1.550617	1.880882
4 11	-0.113598		1.842200	2.387946
5 11	1.073313		0.321815	4.058381
6 11	-0.810886		0.977450	3.199600
7 11	0.774427		1.121489	2.261724
8 11	0.667794	0.401388	0.371076	0.190946
9 11	-4.256761		-1.298245	2.861238
10 11	1.188914	1.040265	1.037336	0.185354
11 11	-0.446754		-0.342607	3.503477
12 11	-0.335805	-1.045615	-0.934479	0.191101
13 11	-1.896477		0.261552	3.302563
14 11	2.228810		0.529430	3.378829
15 11	-0.861164		-0.485036	2.463142
1 12	5.354785	5.210977	5.174006	0.185439
2 12	0.525775		1.746008	1.785812
3 12	0.115824		2.247287	1.908123
4 12	1.771480	2.441880	2.411340	0.191968
5 12	1.081651		1.032614	3.808844
6 12	2.498181		1.677955	2.404052
7 12	3.261475		1.188060	3.490934
8 12	2.201032		0.109920	2.407073
9 12	-2.733236		-1.064478	4.101331
10 12	2.954685		0.909766	2.623859
11 12	0.236108		-0.015879	4.420880
12 12	0.384083		-0.219115	2.204263
13 12	0.204221		0.511039	4.127141
14 12	1.777615		0.627236	3.383824
15 12	-0.693838	-0.623532	-0.586708	0.192586
1 13	4.044871		4.560244	2.442561
2 13	2.860570		2.254862	1.856426
3 13	2.478425	2.684627	2.651045	0.188179
4 13	2.128292		2.557470	2.273103
5 13	-0.226666		1.500788	3.912977

(i,j)	X_{ij}	Y_{ij}	\hat{X}_{ij}	Error est.		
6 13	2.278092	2.397945	2.311546	0.192541		
7 13	2.166856		1.237134	3.696093		
8 13	-0.902468		-0.184998	3.012526		
9 13	-5.633426		-0.901749	4.647931		
10 13	1.880781		0.845099	3.912006		
11 13	0.399469		0.350556	4.684173		
12 13	0.770136		0.569001	2.211693		
13 13	0.594662		0.951726	4.391617		
14 13	0.707030		1.064161	3.462294		
15 13	1.037011		-0.156431	2.570370		
1 14	2.709518		3.989799	3.299059		
2 14	2.163806		2.557711	2.498391	0.190700	
3 14	3.933929			2.478351	1.571898	
4 14	1.047408			2.181163	3.586587	
5 14	1.880644			1.186116	4.691736	
6 14	0.338981	1.666538		2.613564		
7 14	3.702429	0.595144		4.264099		
8 14	-0.950579	-0.900711		2.405537		
9 14	-4.027186	-1.011011		4.730176		
10 14	-0.210480	0.707915		4.496175		
11 14	-0.393181	0.671928		4.379102		
12 14	1.156109	1.422708		1.353979	0.191856	
13 14	-0.853163			1.426693	3.949728	
14 14	1.771989			1.550313	2.470104	
15 14	2.751313			0.259616	3.678948	
1 15	2.849064			3.149082	4.314126	
2 15	4.349098		1.951333	1.968434	2.284982	
3 15	2.273647			1.948093	0.190128	
4 15	1.311287			1.653418	3.858080	
5 15	4.966369			0.780120	5.008107	
6 15	1.530654			1.010959	3.862343	
7 15	4.542632			-0.103889	4.166615	
8 15	-1.694414			-1.868427	-1.779007	0.192609
9 15	-1.921006				-1.355054	4.397062
10 15	0.506533				0.295977	4.902009
11 15	-0.653809				0.457464	4.966026
12 15	-1.218652	1.157454			2.593710	
13 15	-1.411769	1.410477			4.041149	
14 15	1.314279	1.865621			1.814761	0.192803
15 15	0.047197				0.523704	3.921539

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.8$, $\beta = 0.8$ y $\sigma_\eta^2 = 2$, así como de la ecuación de observaciones con todos los $m_{ij} = 1$ y $\sigma_\epsilon^2 = 0.5$.

(i, j)	X_{ij}	Y_{ij}
1 1	8.806044	9.591578
2 1	7.162165	8.074298
3 1	7.942614	7.915698
4 1	3.847925	4.038509
5 1	0.944189	-0.201774
6 1	-2.190621	-2.478594
7 1	-6.369785	-5.577783
8 1	-5.385699	-5.143218
9 1	-1.382817	-1.478414
10 1	-1.521207	-2.934787
11 1	-6.375667	-5.023182
12 1	-2.087674	-2.541390
13 1	-2.178226	-1.562076
14 1	-2.772761	-3.577667
15 1	-0.935388	-0.887629
1 2	5.222718	4.320528
2 2	5.396769	4.988649
3 2	6.528876	5.800175
4 2	5.461758	6.337057
5 2	4.213858	5.005144
6 2	2.222320	2.735626
7 2	1.049303	-0.240267
8 2	3.511186	3.894895
9 2	4.227817	4.717739
10 2	3.481985	2.206802
11 2	-0.415188	-0.805373
12 2	2.891075	2.353479
13 2	0.410058	-0.628020
14 2	0.162853	0.010889
15 2	3.106062	2.472521
1 3	4.158504	5.224587
2 3	4.778273	5.667455
3 3	2.820890	2.628827
4 3	4.265201	6.329114
5 3	1.947418	2.037646
6 3	0.486069	0.546901
7 3	-1.764745	-1.514635
8 3	1.311653	1.203369
9 3	3.931644	3.916668
10 3	4.568302	4.306947

(i, j)	X_{ij}	Y_{ij}
11 3	2.374704	2.144170
12 3	5.634242	6.657089
13 3	4.245383	5.106304
14 3	3.465408	3.990854
15 3	7.490768	7.087829
1 4	6.300088	6.743890
2 4	6.310878	6.300581
3 4	2.973786	3.273618
4 4	4.006611	3.603770
5 4	0.943428	0.379933
6 4	0.177389	0.695617
7 4	-1.600818	-3.010304
8 4	0.874266	1.538326
9 4	5.228074	4.572309
10 4	4.853456	5.191116
11 4	5.092232	5.533254
12 4	7.756491	7.899442
13 4	6.813999	5.499574
14 4	5.592033	5.700435
15 4	9.849261	10.840568
1 5	5.261492	5.765608
2 5	3.897995	4.183728
3 5	4.573069	4.662215
4 5	5.803927	5.505465
5 5	0.925020	0.795079
6 5	-1.519497	-0.198490
7 5	-2.312919	-2.611580
8 5	0.740267	1.066617
9 5	5.827050	5.846249
10 5	5.353712	4.906400
11 5	5.917754	5.738611
12 5	6.403086	6.341691
13 5	5.195774	6.307313
14 5	3.788099	3.757895
15 5	6.871676	7.638011
1 6	7.161623	7.284782
2 6	6.012246	5.660534
3 6	6.805210	6.807968
4 6	7.980737	8.587560
5 6	1.112904	1.014606

(i, j)	X_{ij}	Y_{ij}
6 6	-0.231838	0.540289
7 6	1.297335	0.898997
8 6	3.687971	3.656596
9 6	7.789039	9.123699
10 6	4.578571	4.430729
11 6	4.887241	5.068720
12 6	6.829768	6.909107
13 6	4.928418	4.562776
14 6	4.681614	4.178853
15 6	7.779016	6.702012
1 7	7.387676	7.474910
2 7	5.237883	3.181817
3 7	5.495605	4.128766
4 7	6.542550	6.450309
5 7	0.656283	1.719466
6 7	-0.700631	-1.235443
7 7	2.090572	2.969200
8 7	5.490899	5.802300
9 7	8.102700	8.162654
10 7	3.965196	4.009809
11 7	1.748208	3.170438
12 7	4.482758	4.734808
13 7	2.601280	1.929964
14 7	2.026326	0.226422
15 7	0.828373	1.236706
1 8	8.078638	8.874998
2 8	6.502764	6.620796
3 8	5.490929	5.436166
4 8	5.363171	4.572279
5 8	2.937182	2.336898
6 8	2.389409	2.574546
7 8	4.669699	5.723649
8 8	4.807282	4.639690
9 8	3.453361	3.417566
10 8	0.337209	0.935298
11 8	-0.370091	0.589291
12 8	1.899403	1.409657
13 8	0.782105	1.487240
14 8	-2.306505	-2.056973
15 8	-3.537962	-4.070727
1 9	10.146421	9.903963
2 9	7.508824	8.182357
3 9	6.705804	6.712920
4 9	5.523470	5.430141
5 9	5.041915	4.215514
6 9	4.765445	4.042580
7 9	4.133578	2.757299
8 9	4.171792	4.700425
9 9	0.667092	-0.042671
10 9	0.914875	0.511385

(i, j)	X_{ij}	Y_{ij}
11 9	-0.700235	-0.662158
12 9	3.253816	3.062948
13 9	2.448543	3.172765
14 9	2.671206	3.201169
15 9	0.548501	1.177151
1 10	8.214607	8.550375
2 10	7.188508	7.277879
3 10	6.190748	7.035577
4 10	4.022502	4.976894
5 10	4.035555	4.107299
6 10	0.481103	0.769178
7 10	1.838322	2.626878
8 10	3.962308	5.071205
9 10	-1.475700	-2.268091
10 10	1.846985	1.727105
11 10	0.184885	0.090933
12 10	2.362267	1.751692
13 10	1.251923	1.706589
14 10	4.111719	3.902661
15 10	4.127204	3.252711
1 11	9.533021	10.120527
2 11	6.531045	6.578150
3 11	3.374472	4.492258
4 11	3.297142	3.729393
5 11	3.519573	3.491702
6 11	1.399821	1.902642
7 11	2.331669	2.286690
8 11	2.644788	2.223562
9 11	-2.433188	-3.224648
10 11	0.777952	0.542916
11 11	-0.425873	-0.704736
12 11	0.035336	-1.086970
13 11	-1.847916	-2.209403
14 11	1.525100	0.922679
15 11	-0.274602	-1.469976
1 12	8.294210	8.066829
2 12	5.250651	5.482367
3 12	4.150208	4.029612
4 12	5.154338	6.214333
5 12	4.455787	4.226177
6 12	5.441472	6.244887
7 12	6.916061	7.772117
8 12	6.918368	8.409444
9 12	1.759748	2.326431
10 12	5.489633	6.176147
11 12	3.402244	4.441285
12 12	3.458650	2.983769
13 12	2.669973	3.211335
14 12	4.075429	3.789978
15 12	1.979530	2.090693



(i, j)		X_{ij}	Y_{ij}
1	13	8.887046	8.239456
2	13	8.458028	7.182034
3	13	8.254546	8.580580
4	13	8.033611	7.262882
5	13	5.374948	5.402405
6	13	6.569156	6.758659
7	13	6.884822	6.662981
8	13	4.160767	4.540993
9	13	-2.381207	-1.702110
10	13	2.279494	1.972169
11	13	1.371744	2.416990
12	13	2.091117	3.039768
13	13	1.868106	1.497002
14	13	2.265342	2.672327
15	13	2.564175	2.929966
1	14	9.249210	9.769529
2	14	8.609907	9.232725
3	14	10.147102	9.994055
4	14	7.986562	9.239203
5	14	7.790424	8.496733
6	14	5.960384	6.404234
7	14	8.628547	7.789652
8	14	4.824387	5.281613
9	14	-0.414658	-0.029414
10	14	1.325515	1.188413
11	14	0.573521	-0.574634
12	14	2.074776	2.496305
13	14	0.314353	0.646253
14	14	2.564590	3.255985
15	14	4.192148	3.683337
1	15	10.831720	10.459784
2	15	12.074127	10.991372
3	15	10.635658	10.126034
4	15	9.166241	8.268649
5	15	11.825813	12.894879
6	15	9.274945	8.972935
7	15	11.827212	12.384360
8	15	6.199495	5.924357
9	15	2.983162	2.582341
10	15	3.991933	4.409868
11	15	2.212573	2.103829
12	15	1.102713	2.361595
13	15	-0.236092	0.637877
14	15	1.917063	2.788811
15	15	1.234893	1.170476

Datos simulados de un campo espacial $AR(1) \times AR(1)$ con parámetros $\lambda = 0.3$, $\beta = 0.8$ y $\sigma_\eta^2 = 3$, datos Y_{ij} de la ecuación de observación las constantes m_{ij} generadas según una distribución uniforme (0,10) y $\sigma_\epsilon^2 = 0.7$, datos de X alisados y errores esperados.

(i,j)	X_{ij}	Y_{ij}	m_{ij}	\hat{X}_{ij}	Error est.
1 1	4.649611	28.711746	5.975184	4.790378	0.014029
2 1	1.538583	8.669324	4.933157	1.763463	0.020584
3 1	3.171790	4.991205	1.583664	3.065121	0.188398
4 1	-2.117876	-13.856855	6.649281	-2.074807	0.011334
5 1	-3.249153	-15.928393	4.485005	-3.535111	0.024799
6 1	-4.582811	-43.836805	9.491132	-4.615510	0.005573
7 1	-7.029844	-51.580165	7.470616	-6.900876	0.008992
8 1	-2.463971	-21.051259	8.660071	-2.421524	0.006692
9 1	2.844096	27.826133	9.823593	2.829151	0.005203
10 1	0.345016	0.100535	5.139194	0.019389	0.018925
11 1	-6.214587	-25.996043	4.440572	-5.834528	0.025294
12 1	1.825607	4.389860	2.698665	1.636324	0.067512
13 1	-0.074594	0.246823	6.464484	0.030864	0.011988
14 1	-1.284086	-11.985261	8.592008	-1.392985	0.006798
15 1	1.185902	7.040827	5.889452	1.200122	0.014459
1 2	1.818336	6.268840	4.034635	1.574099	0.030434
2 2	2.037968	-0.232675	0.122778	1.880273	1.651628
3 2	3.319866	10.882462	3.537694	3.039485	0.039357
4 2	1.288253	11.375603	8.026320	1.417613	0.007773
5 2	0.195969	2.577547	8.375222	0.300829	0.007146
6 2	-1.348154	-2.589207	2.371063	-1.145157	0.085362
7 2	-1.296738	-2.122950	0.460470	-4.032381	1.012950
8 2	2.883183	26.758441	9.123399	2.922666	0.006022
9 2	2.602706	18.728454	6.973038	2.694034	0.010283
10 2	0.902957	3.779720	5.856905	0.649527	0.014540
11 2	-3.649247	-26.082342	7.020808	-3.714255	0.010145
12 2	2.852854	24.283636	8.735014	2.779253	0.006568
13 2	-1.474589	-15.068543	9.385844	-1.598743	0.005691
14 2	-0.644696	-5.264660	7.887203	-0.670055	0.008050
15 2	3.451162	0.083426	0.241380	2.875862	1.537146
1 3	0.708394	6.134588	6.879195	0.898028	0.010570
2 3	1.990199	18.244640	8.638605	2.109180	0.006720
3 3	-0.629802	-5.927381	9.050667	-0.648893	0.006119
4 3	2.270945	12.809784	4.565379	2.782384	0.023818
5 3	-1.112651	-0.274675	0.342815	-0.818908	1.243388
6 3	-1.646558	-2.731084	1.702376	-1.553768	0.158769
7 3	-3.131579	-5.481115	1.844771	-3.012653	0.137748
8 3	2.396057	12.053251	5.083923	2.372647	0.019248
9 3	4.248925	23.399604	5.511351	4.249185	0.016403
10 3	3.017474	27.737233	9.294684	2.980028	0.005803

(i,j)	X_{ij}	Y_{ij}	m_{ij}	\hat{X}_{ij}	Error est.
11 3	-0.662354	-4.096736	5.773287	-0.703916	0.014962
12 3	4.375077	8.403839	1.644220	5.022345	0.167986
13 3	0.991626	5.391792	4.410062	1.213307	0.025484
14 3	0.382120	4.416315	9.930368	0.445211	0.005086
15 3	5.893523	56.777555	9.714786	5.842153	0.005317
1 4	2.333026	15.513116	6.424274	2.415351	0.012109
2 4	2.256322	6.275258	2.786588	2.203445	0.062470
3 4	-1.864345	-7.458421	4.190847	-1.743467	0.028182
4 4	1.434069	7.531422	5.584158	1.356450	0.016020
5 4	-2.339981	-7.595924	2.961214	-2.556777	0.055868
6 4	-1.409104	-12.266888	9.140600	-1.346492	0.006000
7 4	-2.557130	-17.211829	6.078727	-2.829277	0.013508
8 4	1.872088	10.471959	5.174025	2.029130	0.018588
9 4	6.108082	59.342440	9.842427	6.027758	0.005177
10 4	2.654224	4.847367	1.675759	2.937333	0.162319
11 4	2.277556	10.765106	4.497487	2.378591	0.024521
12 4	5.193641	48.377585	9.282204	5.210168	0.005820
13 4	2.303725	12.295005	6.012112	2.041717	0.013805
14 4	0.863603	4.944643	5.577075	0.884900	0.016024
15 4	6.842861	27.896974	3.905390	7.125527	0.032497
1 5	3.701741	29.810905	7.892076	3.777347	0.008041
2 5	0.729383	1.893677	2.132751	0.978368	0.103998
3 5	2.000418	10.420009	5.156187	2.003317	0.018735
4 5	3.227781	-0.225647	0.039500	3.120330	1.742737
5 5	-3.585416	-14.061761	3.879050	-3.634614	0.032859
6 5	-3.842952	-18.456391	5.209386	-3.528462	0.018357
7 5	-2.496825	-10.760505	4.168143	-2.563710	0.028480
8 5	2.423780	10.044976	3.985029	2.519831	0.031105
9 5	7.138473	45.616272	6.387017	7.140284	0.012243
10 5	2.989153	19.235984	6.612323	2.902132	0.011430
11 5	2.898940	9.417987	3.321886	2.813333	0.044440
12 5	2.913638	2.702717	0.952541	3.309827	0.423031
13 5	0.963871	10.347383	9.370747	1.102414	0.005710
14 5	-0.162181	-0.710830	4.162573	-0.152916	0.028578
15 5	4.655831	2.675999	0.380009	6.125061	1.223189
1 6	5.877121	40.215294	6.817890	5.895431	0.010759
2 6	2.109675	17.055658	8.281753	2.056346	0.007303
3 6	3.076774	4.395527	1.427554	2.976081	0.216843
4 6	4.029682	12.452327	2.911972	4.235695	0.057847
5 6	-5.247564	-8.096987	1.520834	-5.059103	0.195259
6 6	-2.948631	-1.071658	0.673279	-2.296539	0.679056
7 6	0.931470	4.935559	5.804673	0.843881	0.014804
8 6	3.525140	32.224396	9.151838	3.524354	0.005985
9 6	6.983669	36.223726	4.960792	7.283694	0.020237
10 6	0.071013	0.254369	6.045339	0.053081	0.013656
11 6	1.520863	6.335596	4.024601	1.555799	0.030516
12 6	4.032483	6.056339	1.478608	4.023761	0.204982
13 6	0.554020	0.105446	0.971227	0.311216	0.408794
14 6	1.071145	3.071225	3.422597	0.875387	0.041940
15 6	5.261627	17.680737	3.602509	4.869883	0.038099

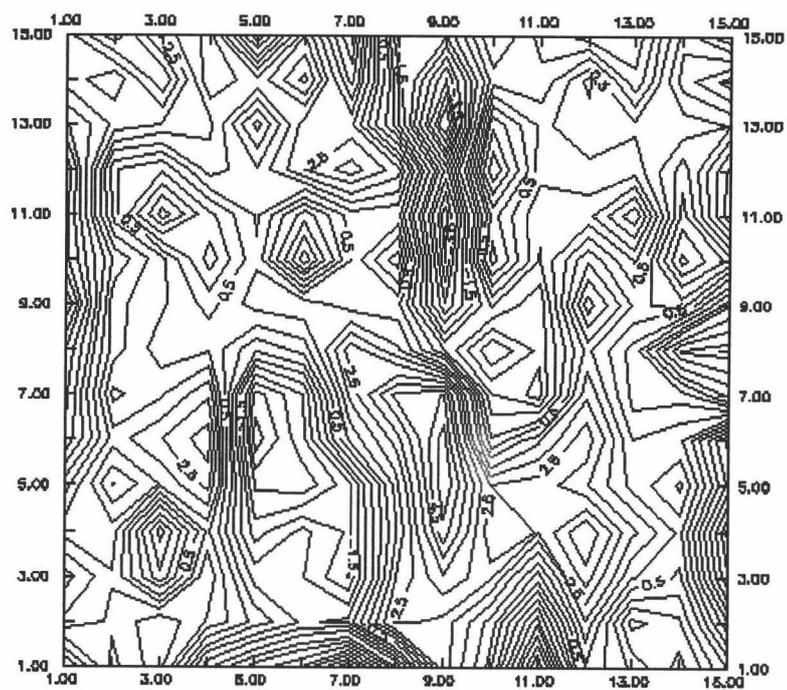
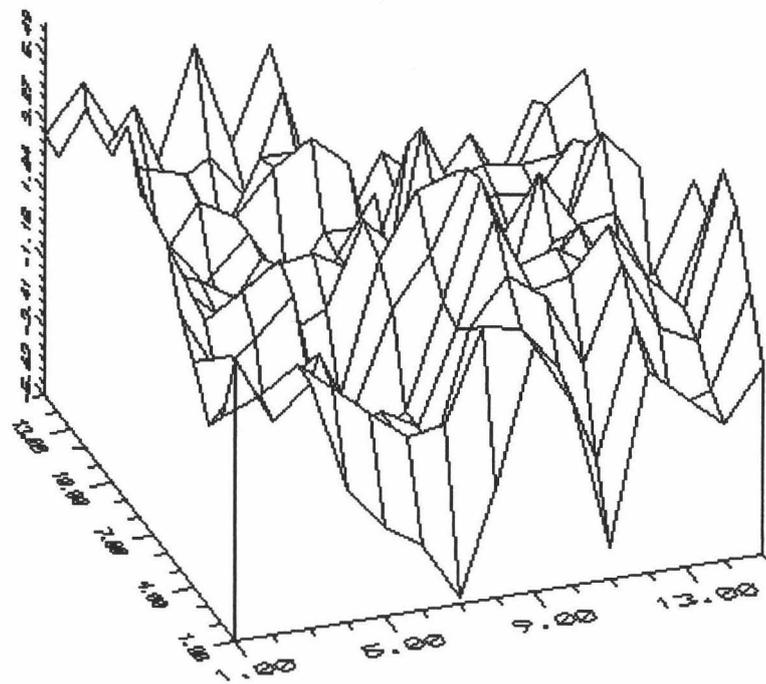
(i,j)	X_{ij}	Y_{ij}	m_{ij}	\hat{X}_{ij}	Error est.
1 7	6.299687	46.631813	7.385858	6.312760	0.009176
2 7	1.066562	1.963211	4.121635	0.495878	0.029169
3 7	1.918626	2.842171	2.324286	1.246705	0.088525
4 7	3.203971	13.607112	4.281016	3.167937	0.027028
5 7	-4.645392	-3.594870	1.044657	-3.452920	0.366456
6 7	-2.894736	-22.518443	7.560496	-2.969615	0.008760
7 7	2.378472	23.085307	9.268847	2.490194	0.005835
8 7	5.390158	8.546670	1.517249	5.373425	0.194941
9 7	6.160826	2.634494	0.416105	5.095282	1.094726
10 7	-1.234389	-4.254385	3.489312	-1.223100	0.040403
11 7	-2.114290	-8.624210	4.874928	-1.743924	0.020921
12 7	3.143060	9.515510	2.932581	3.216667	0.056637
13 7	-0.263365	-2.872973	7.892693	-0.361329	0.008039
14 7	-0.146001	-2.496030	2.509264	-1.034380	0.076396
15 7	-1.014639	-2.781172	3.217218	-0.819871	0.047374
1 8	6.765327	13.119038	1.799879	7.312092	0.143752
2 8	2.078408	1.324268	0.569960	1.742348	0.820275
3 8	0.977129	9.045404	9.323437	0.971373	0.005768
4 8	1.481665	12.422519	9.015744	1.379260	0.006166
5 8	-1.213014	-9.952214	7.618990	-1.305735	0.008624
6 8	-0.315327	-0.532439	2.383230	-0.236455	0.084276
7 8	3.283458	17.529845	4.959037	3.509878	0.020216
8 8	2.297379	14.851930	6.551040	2.276595	0.011644
9 8	0.208544	0.655638	3.346960	0.198556	0.043942
10 8	-2.908029	-6.194513	2.373490	-2.528337	0.085100
11 8	-1.656073	-0.946631	1.257063	-1.036355	0.269191
12 8	2.192077	15.788831	7.467030	2.121990	0.008976
13 8	-0.245525	-1.226143	8.392097	-0.149292	0.007113
14 8	-3.664842	-21.642806	5.986084	-3.591002	0.013925
15 8	-3.172649	-26.414303	8.126939	-3.245867	0.007585
1 9	8.781283	82.917348	9.475178	8.746239	0.005586
2 9	1.889357	18.419409	9.327231	1.977217	0.005765
3 9	1.422590	3.971004	2.785470	1.419217	0.062514
4 9	0.621299	3.250210	5.409048	0.602545	0.017023
5 9	0.949576	8.408554	9.884794	0.849499	0.005132
6 9	1.181279	3.577987	3.752958	0.895772	0.035004
7 9	0.747799	2.834063	5.967509	0.493750	0.014025
8 9	1.283657	8.243627	5.934713	1.395110	0.014165
9 9	-2.885389	-14.485048	4.729083	-3.059636	0.022206
10 9	-0.398742	-1.156071	1.701989	-0.674585	0.158159
11 9	-1.873624	-9.959298	5.339572	-1.852561	0.017473
12 9	4.109095	8.772249	2.189797	3.901718	0.098987
13 9	1.043493	4.951160	3.923596	1.256338	0.032099
14 9	1.185521	5.240555	3.891531	1.318985	0.032683
15 9	-1.589805	-7.162629	4.973222	-1.426762	0.020121
1 10	7.249512	36.341356	4.958136	7.341508	0.020241
2 10	2.930303	4.195019	1.395511	2.835199	0.224755
3 10	1.417907	7.190703	4.366355	1.622410	0.025994
4 10	-0.713757	-2.678803	5.335219	-0.488524	0.017493
5 10	0.787166	7.197799	9.036097	0.793501	0.006138

(i,j)	X_{ij}	Y_{ij}	m_{ij}	\hat{X}_{ij}	Error est.
6 10	-3.128641	-29.990696	9.694799	-3.083909	0.005335
7 10	0.841500	1.350878	0.496548	1.564408	0.937314
8 10	3.304086	19.478340	5.498123	3.521460	0.016487
9 10	-4.698384	-33.679041	6.968665	-4.822570	0.010296
10 10	2.298456	5.292093	2.364168	2.148871	0.085672
11 10	-0.893695	-4.199846	4.575029	-0.916293	0.023706
12 10	2.443917	5.430055	2.517474	2.155752	0.075996
13 10	-0.048078	0.480897	1.187030	0.315518	0.298447
14 10	3.794754	1.350264	0.421008	3.152653	1.086028
15 10	2.164552	11.731999	5.898084	1.966734	0.014353
1 11	7.397397	68.123126	9.115094	7.472741	0.006034
2 11	0.877667	6.551640	7.401325	0.886282	0.009135
3 11	-2.002923	-6.851945	4.081299	-1.640095	0.029693
4 11	0.130987	1.087179	4.395337	0.241261	0.025745
5 11	1.119348	2.691914	2.434354	1.074497	0.081167
6 11	-1.398234	-5.280588	4.202109	-1.243338	0.028038
7 11	1.064690	5.119727	4.858638	1.064388	0.021083
8 11	1.274037	11.142769	9.137226	1.224252	0.006004
9 11	-5.189177	-49.577893	9.373629	-5.288534	0.005705
10 11	1.780068	4.348638	2.599190	1.709866	0.071474
11 11	-0.749799	-3.776821	4.597048	-0.821519	0.023567
12 11	0.235607	-0.716196	2.596423	-0.247602	0.071862
13 11	-2.227166	-18.430380	8.083214	-2.270290	0.007666
14 11	3.010289	13.083890	4.583174	2.836962	0.023672
15 11	-0.927518	-10.151832	9.420238	-1.072471	0.005651
1 12	7.758304	45.862804	5.946124	7.702150	0.014119
2 12	0.631565	4.391667	6.519505	0.683618	0.011754
3 12	0.127850	0.283146	3.330755	0.090479	0.044265
4 12	2.284747	1.289679	0.015527	1.351861	1.769780
5 12	1.092426	0.785977	0.968169	0.835724	0.417679
6 12	2.626382	6.283301	2.030431	2.984449	0.114130
7 12	3.926793	22.434167	5.455156	4.103937	0.016741
8 12	2.874945	15.602675	4.813451	3.221570	0.021449
9 12	-3.760862	-35.751017	9.684354	-3.695409	0.005346
10 12	3.870947	19.970115	4.949129	4.023573	0.020313
11 12	-0.050555	1.228115	0.025625	0.803684	1.962131
12 12	0.887292	0.050789	0.690500	0.443684	0.675137
13 12	0.147452	1.413582	5.242592	0.257815	0.018116
14 12	2.419568	5.101596	2.248064	2.258311	0.094135
15 12	-0.842799	-2.589310	3.228338	-0.776956	0.047052
1 13	6.074106	51.982425	8.684185	5.986992	0.006646
2 13	3.473667	16.224939	5.105473	3.167373	0.019092
3 13	2.864671	22.418153	7.691068	2.914284	0.008463
4 13	2.610755	8.972165	3.785918	2.352490	0.034599
5 13	-0.505132	-4.982323	9.927708	-0.494428	0.005089

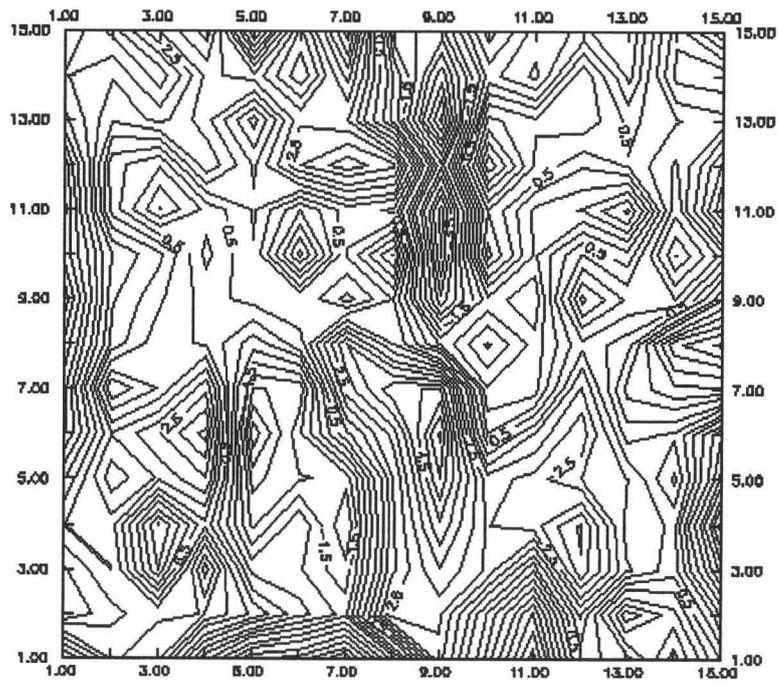
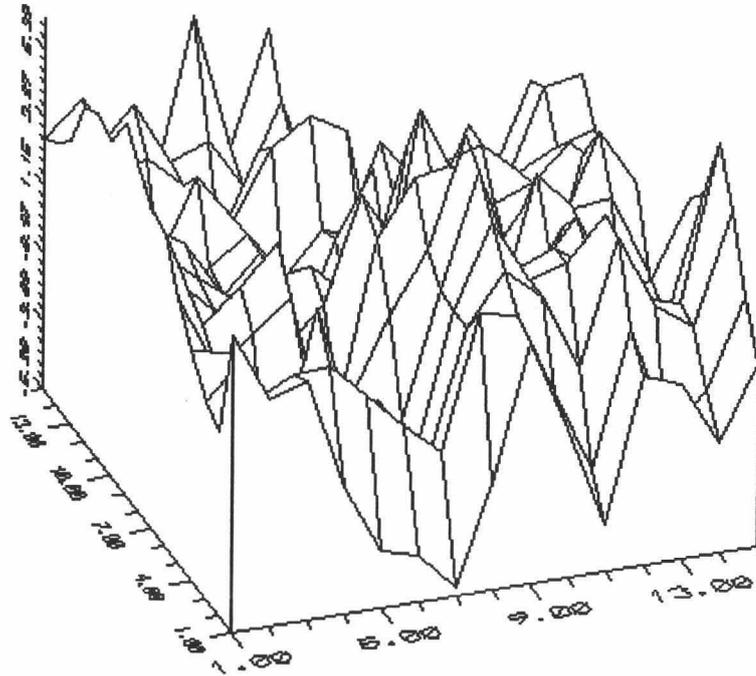
(i,j)	X_{ij}	Y_{ij}	m_{ij}	\hat{X}_{ij}	Error est.
6 13	2.627649	13.366993	5.001721	2.649536	0.019879
7 13	2.784012	10.708473	3.940700	2.743093	0.031805
8 13	-0.814638	-0.651660	1.352196	-0.394052	0.238524
9 13	-7.237465	-45.245958	6.362652	-7.094120	0.012339
10 13	2.953656	20.598708	7.097082	2.896605	0.009932
11 13	0.332694	1.458003	0.665028	1.383452	0.764239
12 13	1.316864	10.514821	7.132367	1.476060	0.009838
13 13	0.634145	1.906509	3.698845	0.492354	0.036009
14 13	1.134346	7.853608	6.498947	1.218334	0.011828
15 13	1.261192	10.292608	7.817840	1.314601	0.008194
1 14	4.276337	19.653006	4.451790	4.420096	0.025045
2 14	2.765503	4.165420	1.239735	3.544458	0.275320
3 14	4.821310	29.876741	6.234369	4.778050	0.012848
4 14	1.285804	2.832620	1.050294	2.553707	0.366350
5 14	2.101823	5.682633	2.306053	2.480846	0.089970
6 14	0.297471	2.857296	7.839840	0.372413	0.008146
7 14	4.817050	19.294045	4.211423	4.566561	0.028024
8 14	-1.100456	-0.993218	1.394163	-0.674118	0.236540
9 14	-5.564902	-9.001113	1.699389	-5.365614	0.161082
10 14	0.360227	0.431097	1.647065	0.411757	0.168379
11 14	-0.488248	-2.449165	2.233796	-1.024578	0.095859
12 14	1.832663	6.772366	3.423218	1.928016	0.041898
13 14	-1.098055	-4.044026	4.040538	-0.980535	0.030275
14 14	2.503549	24.185964	9.333904	2.590132	0.005754
15 14	3.372601	16.058102	4.939846	3.226265	0.020388
1 15	4.224178	16.478736	4.005232	4.117198	0.031058
2 15	5.442103	31.075183	5.945552	5.214983	0.014166
3 15	2.828419	18.945519	6.911463	2.746786	0.010494
4 15	1.654058	0.534355	0.965141	0.928052	0.464256
5 15	5.998776	8.113945	1.141734	6.536193	0.343207
6 15	1.572190	14.002594	9.133713	1.533893	0.006018
7 15	5.869421	2.604482	0.331422	5.633695	1.881402
8 15	-2.234627	-0.794955	0.210060	-0.799621	2.618716
9 15	-3.091015	-1.989956	0.490356	-4.201580	1.283432
10 15	1.038904	1.964237	1.414691	1.315804	0.233128
11 15	-0.889770	-6.104817	6.716504	-0.910204	0.011110
12 15	-1.084258	-3.164023	4.291923	-0.717119	0.027062
13 15	-1.694864	-6.364108	4.365069	-1.457804	0.026170
14 15	2.070776	8.762550	3.733422	2.335658	0.035658
15 15	0.255330	1.874742	7.640909	0.252139	0.008592

Apéndice B

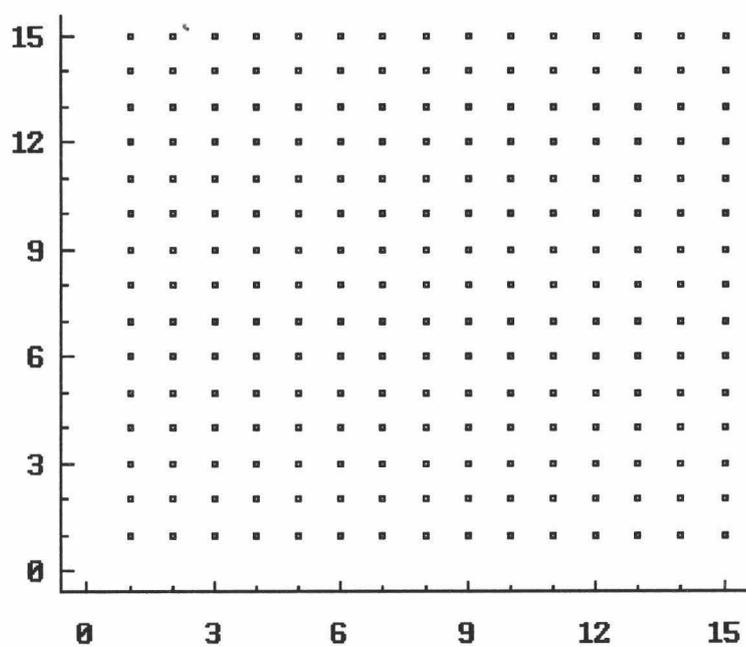
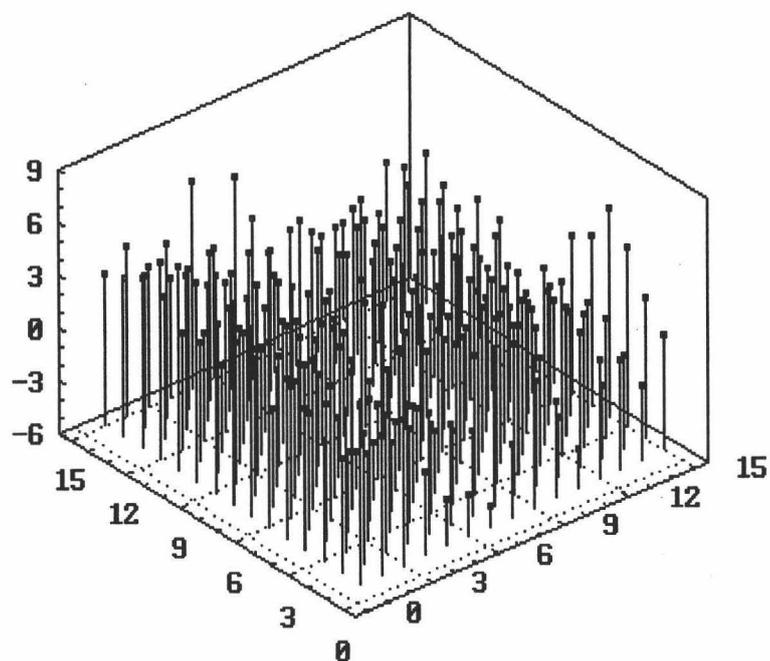
Gráficos



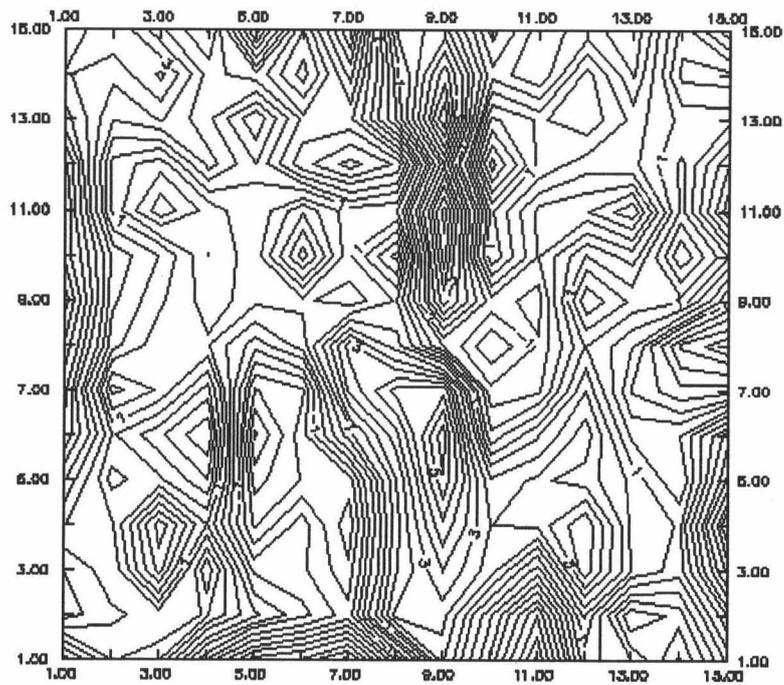
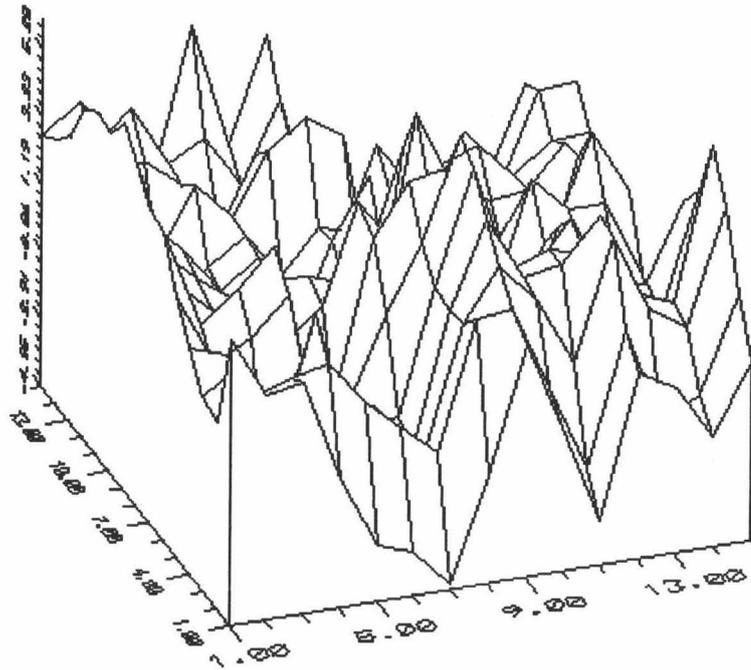
Datos X_{ij} originales primera simulación.



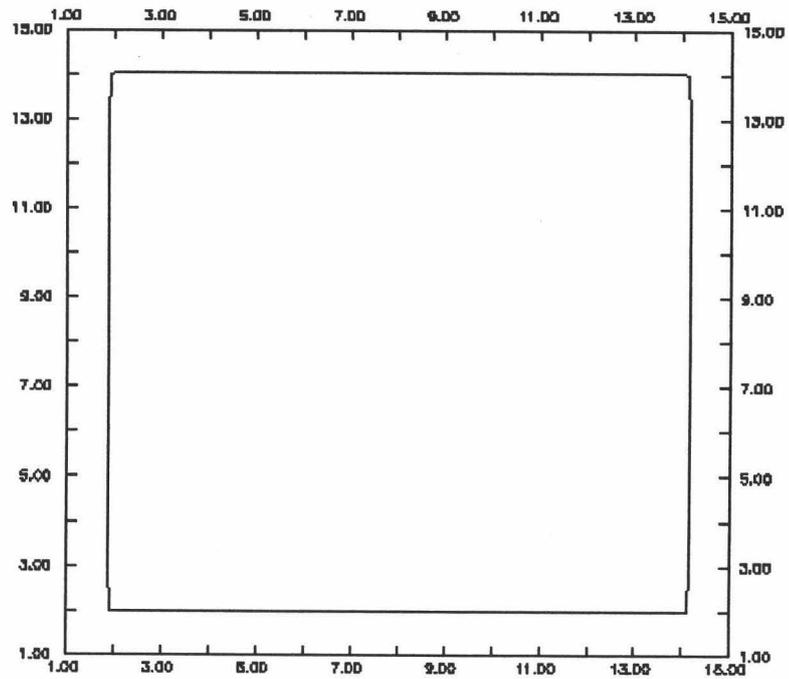
Datos de Y_{ij} .



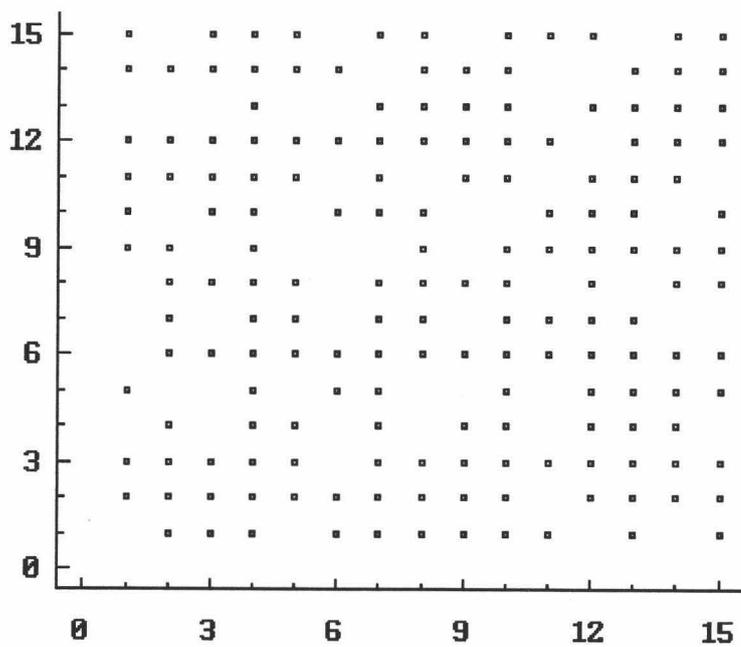
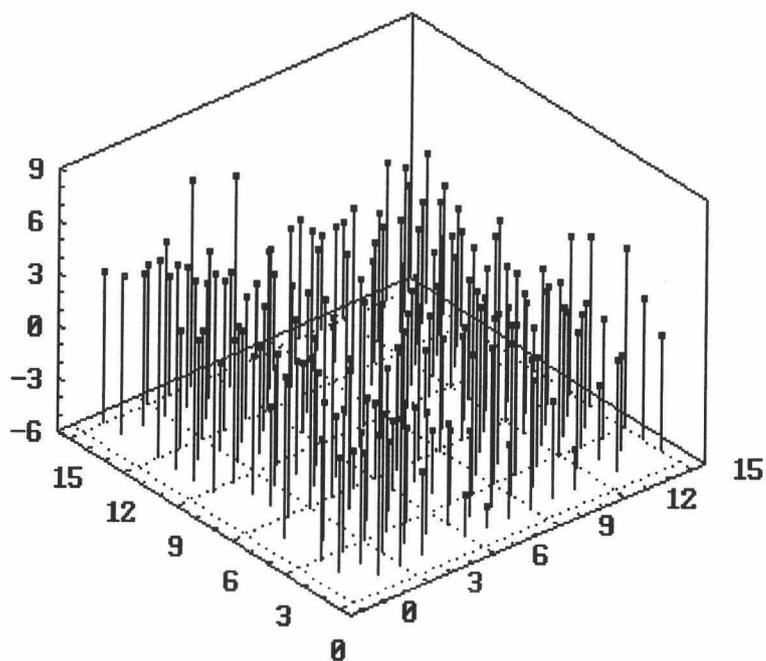
Datos de Y_{ij} observados (100%).



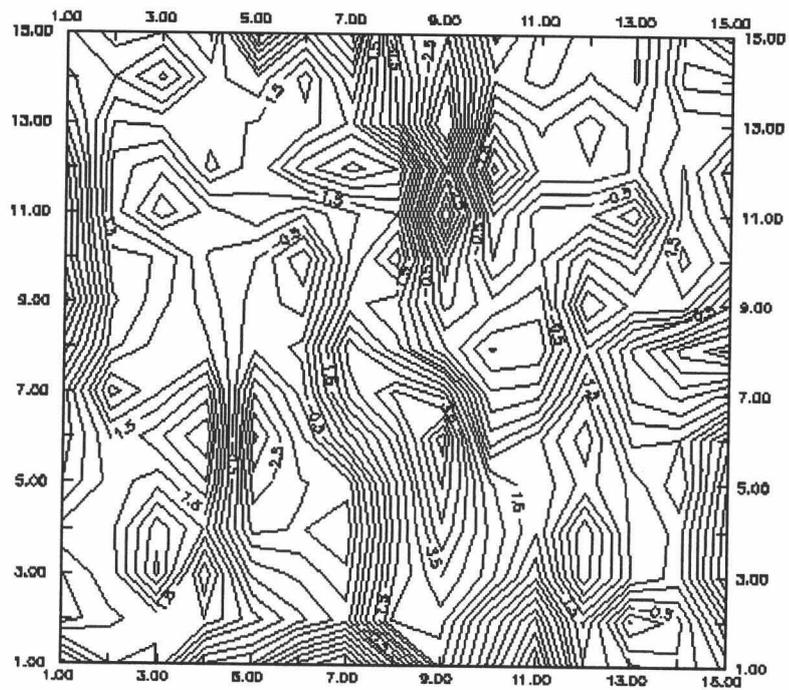
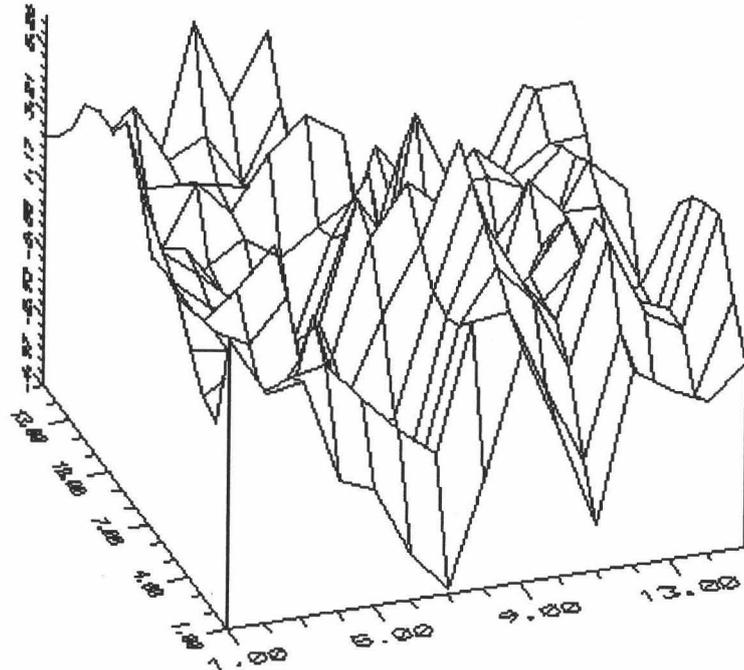
Datos alisados de X con el 100%.



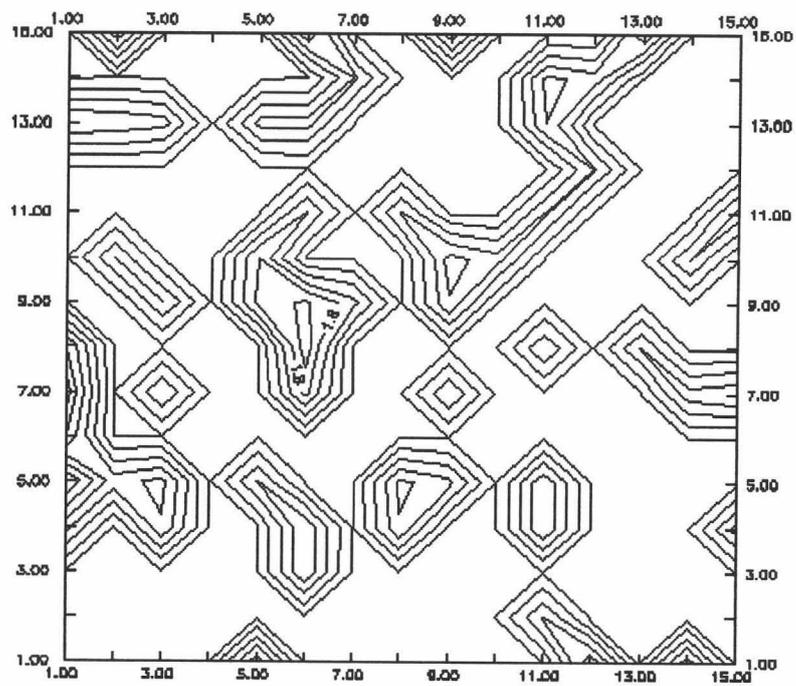
Error cuadrático medio estimado con el 100%.



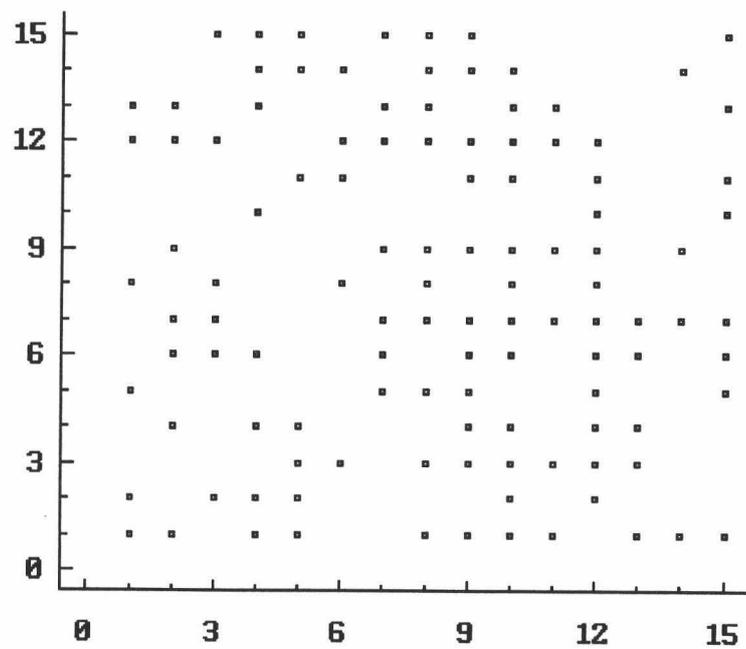
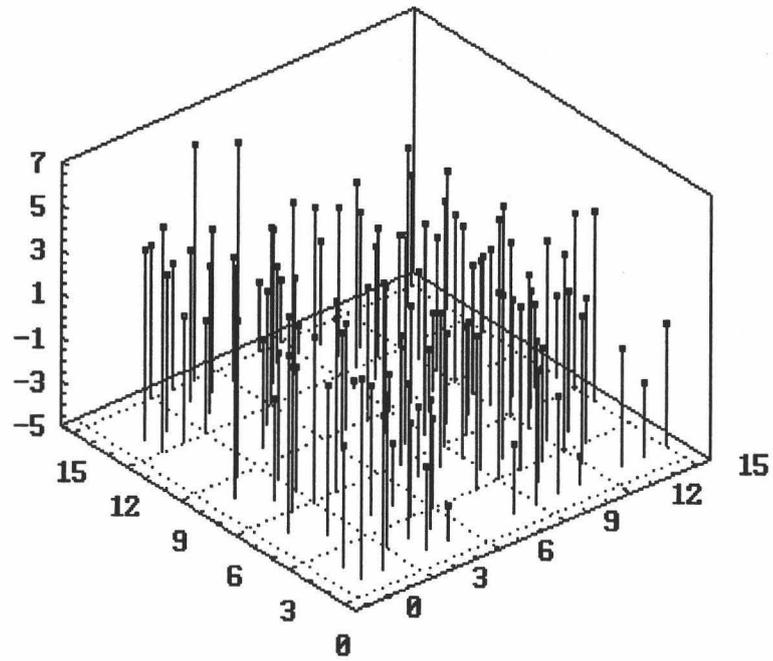
Datos de Y_{ij} observados (75%).



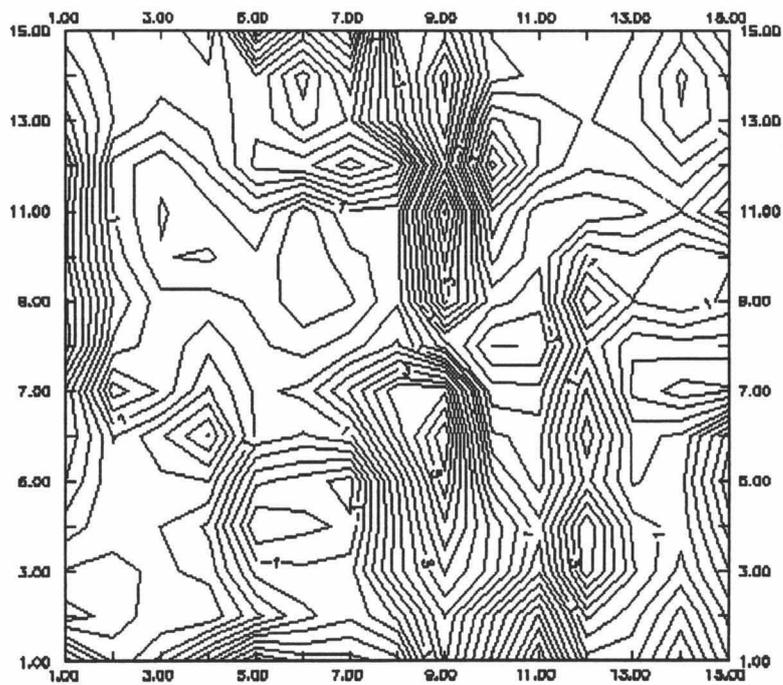
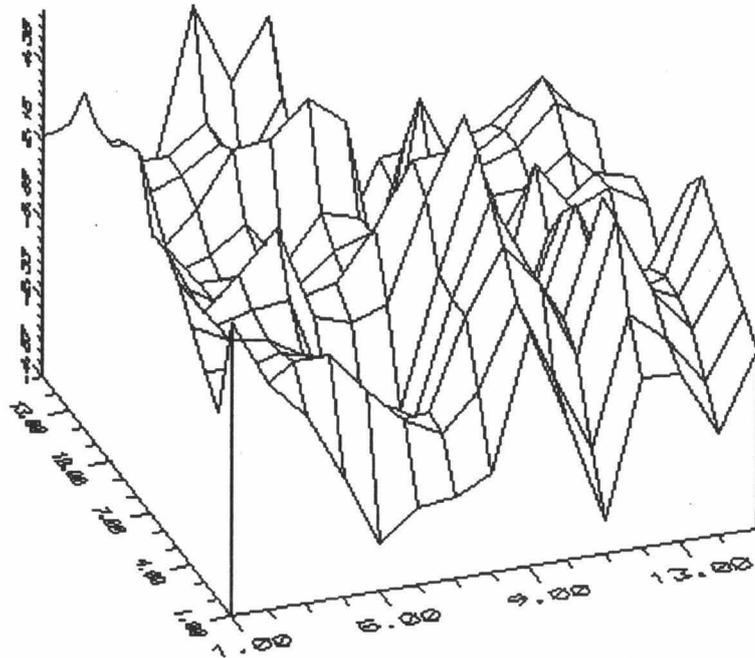
Datos \hat{X}_{ij} en base al 75%.



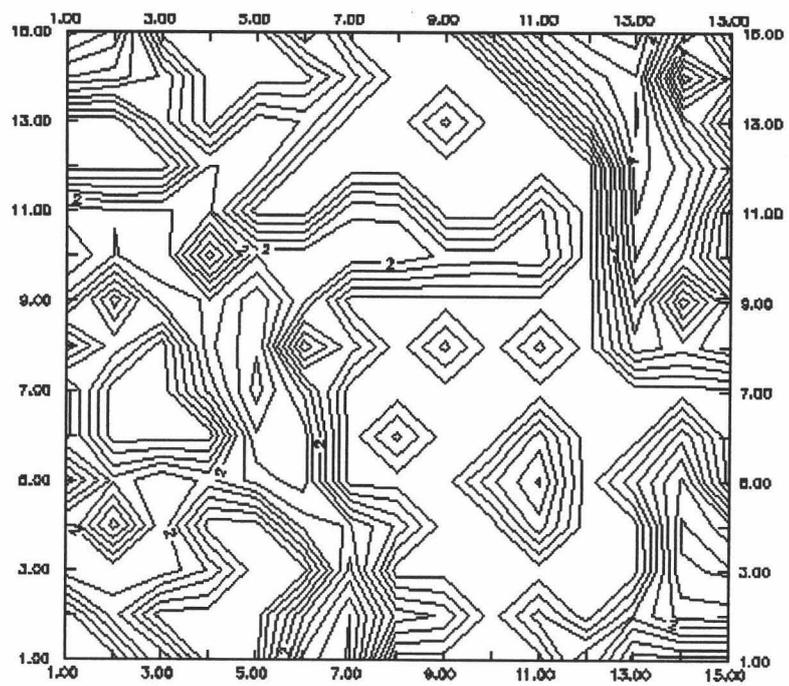
Error cuadrático medio con el 75%.



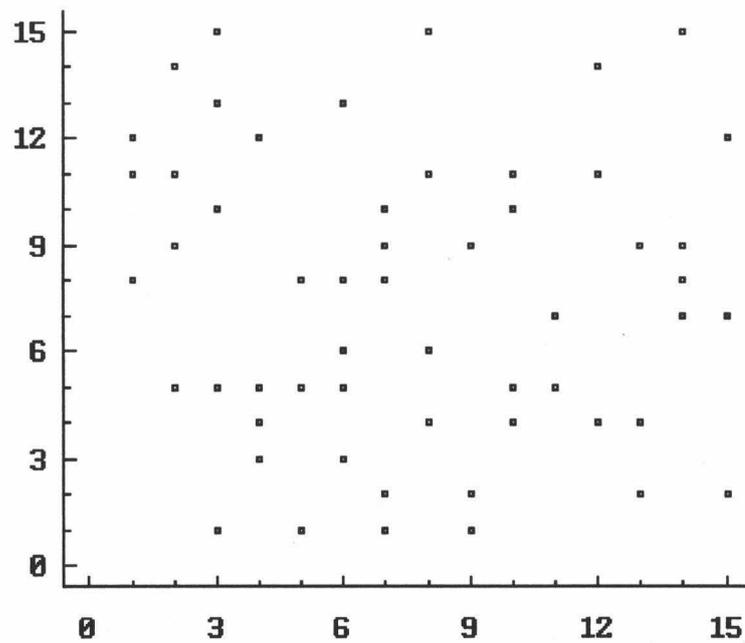
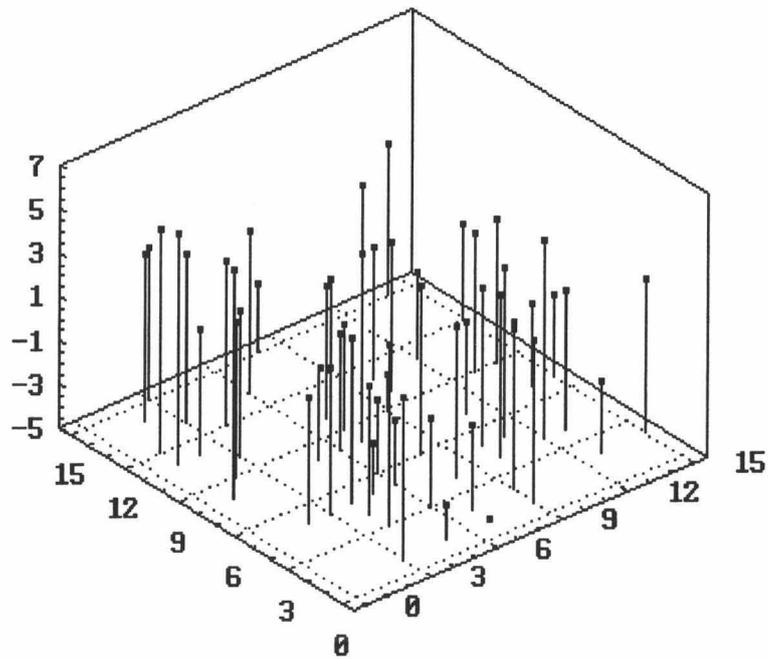
Datos de Y_{ij} observados (50%).



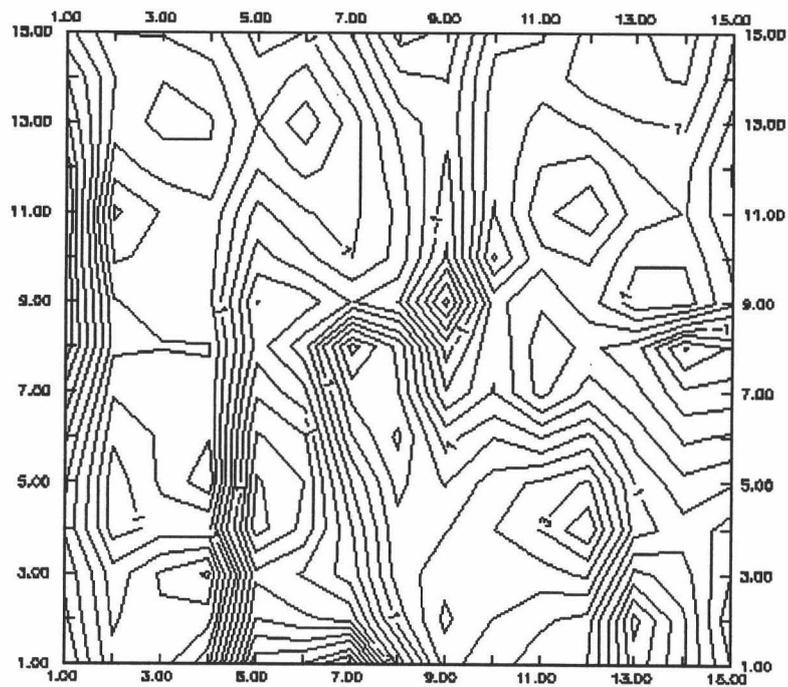
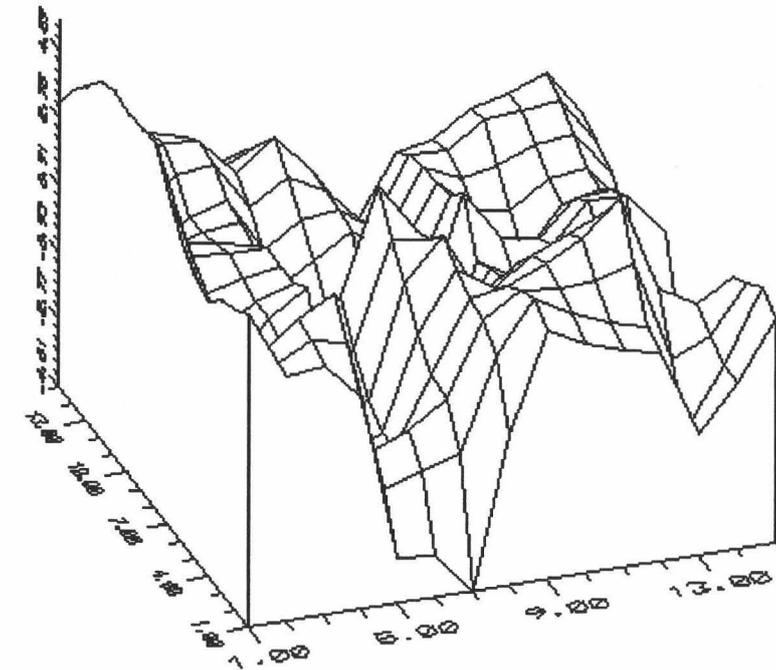
Datos \hat{X}_{ij} en base al 50%.



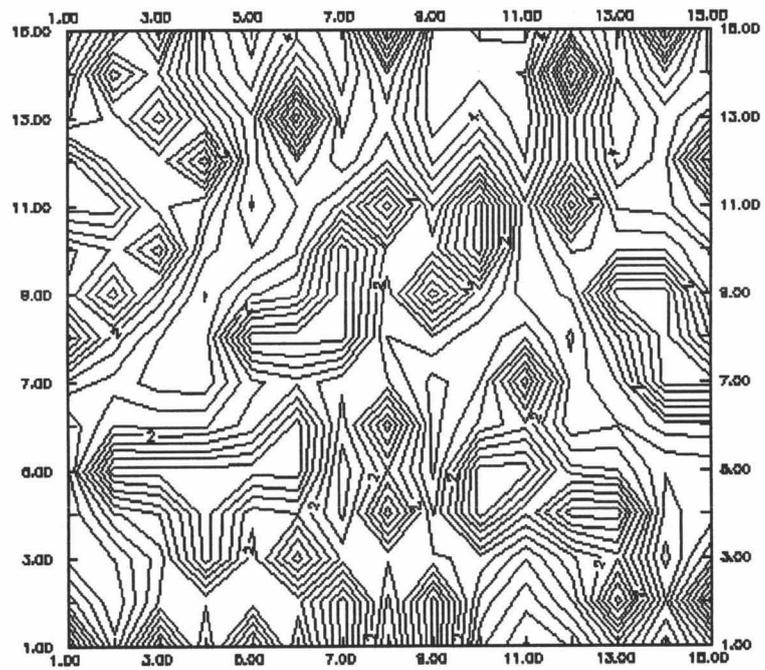
Error cuadrático medio con el 50%.



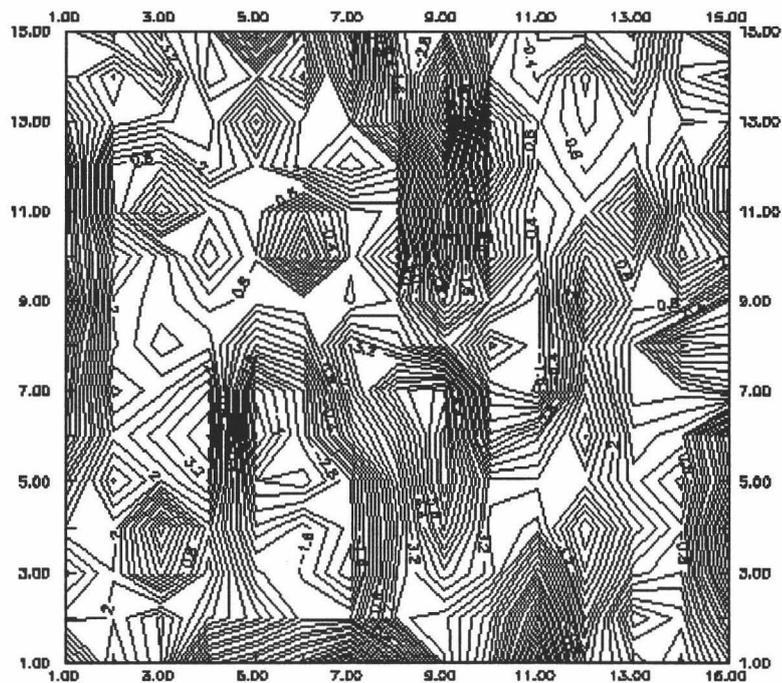
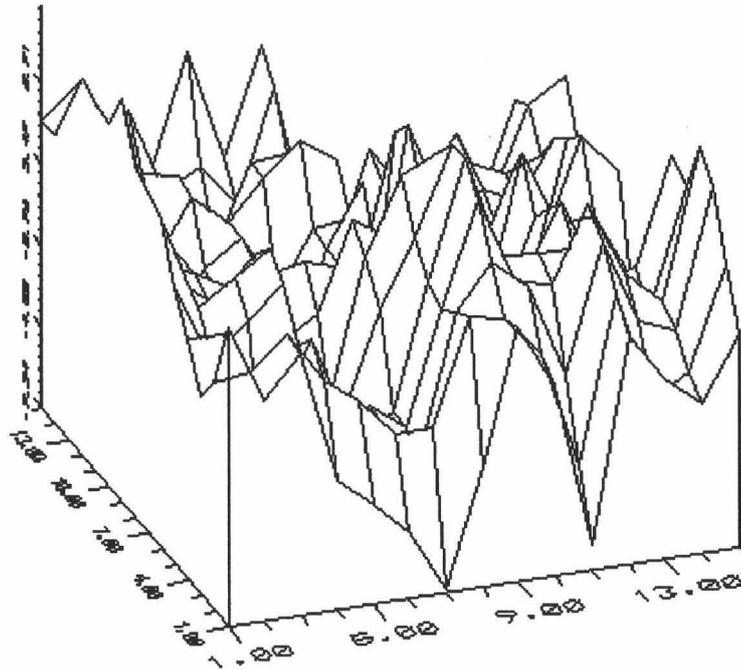
Datos de Y_{ij} observados (25%).



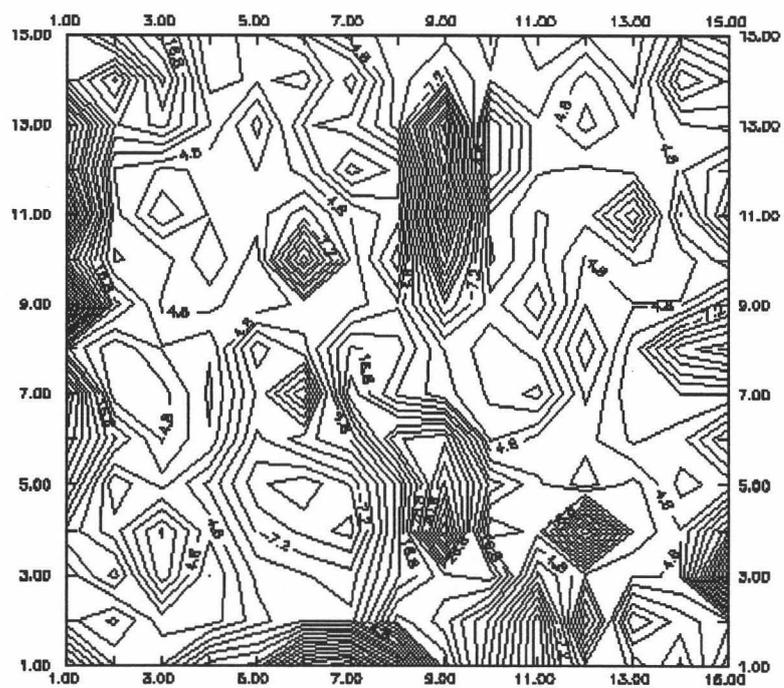
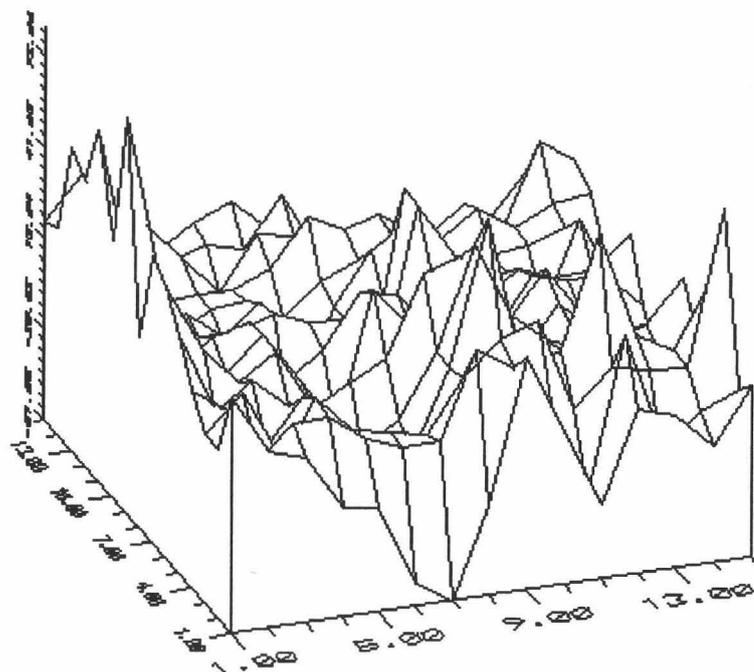
Datos \hat{X}_{ij} en base al 25%.



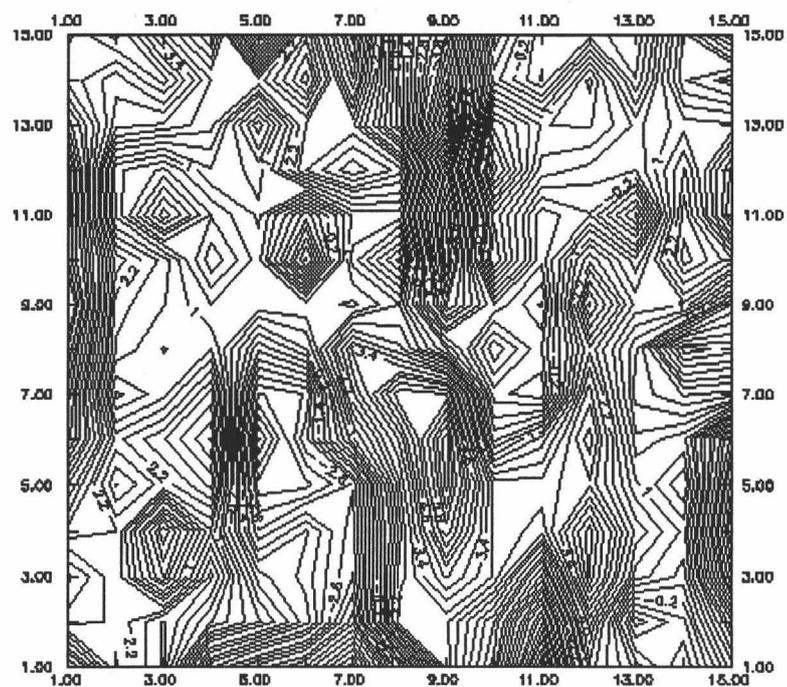
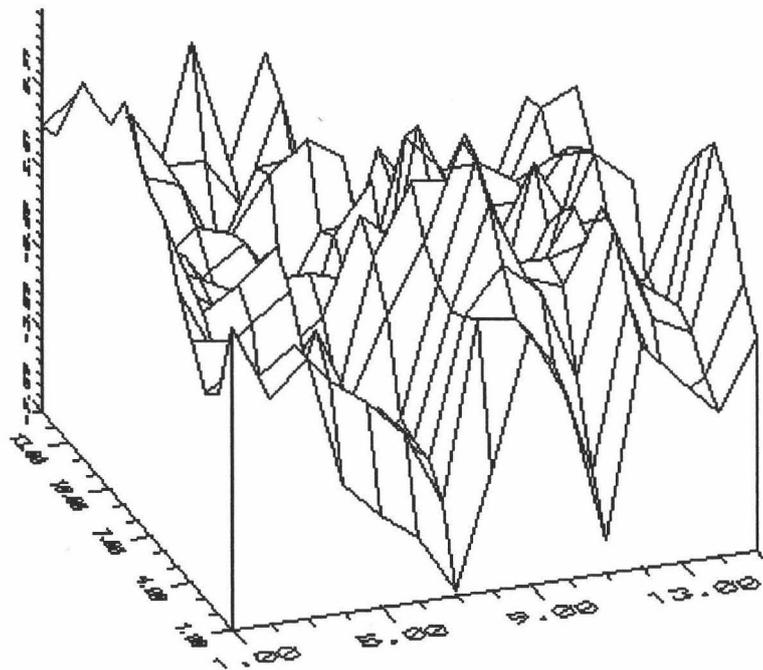
Error cuadrático medio con el 25%.



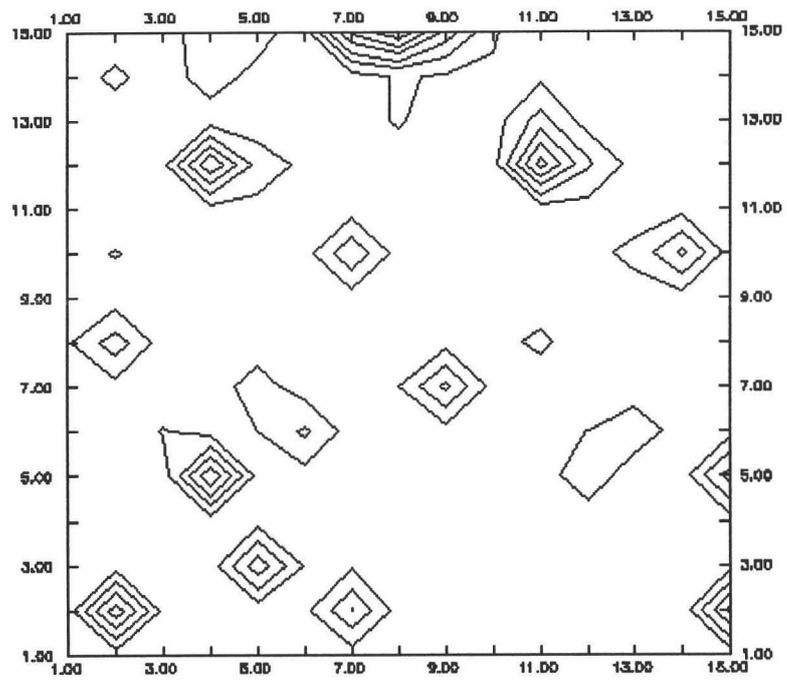
Datos X_{ij} (m_{ij} uniformes).



Datos Y_{ij} observados (m_{ij} uniformes).



Datos \hat{X}_{ij} .



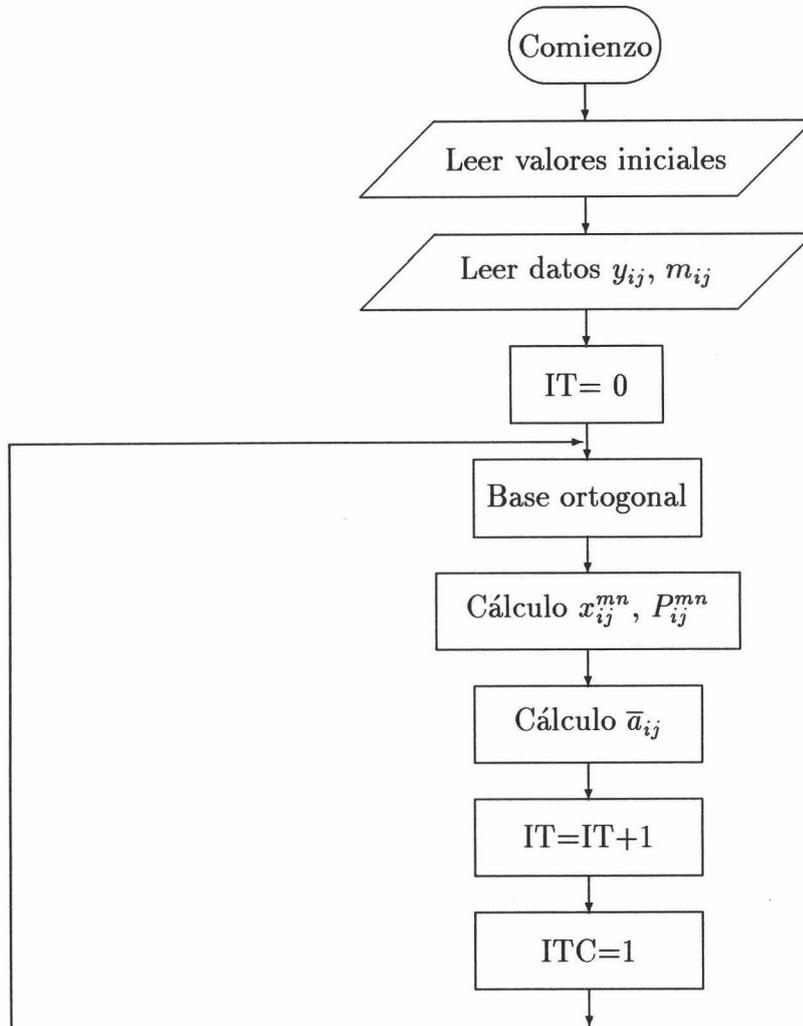
Error cuadrático medio, m_{ij} aleatorias.

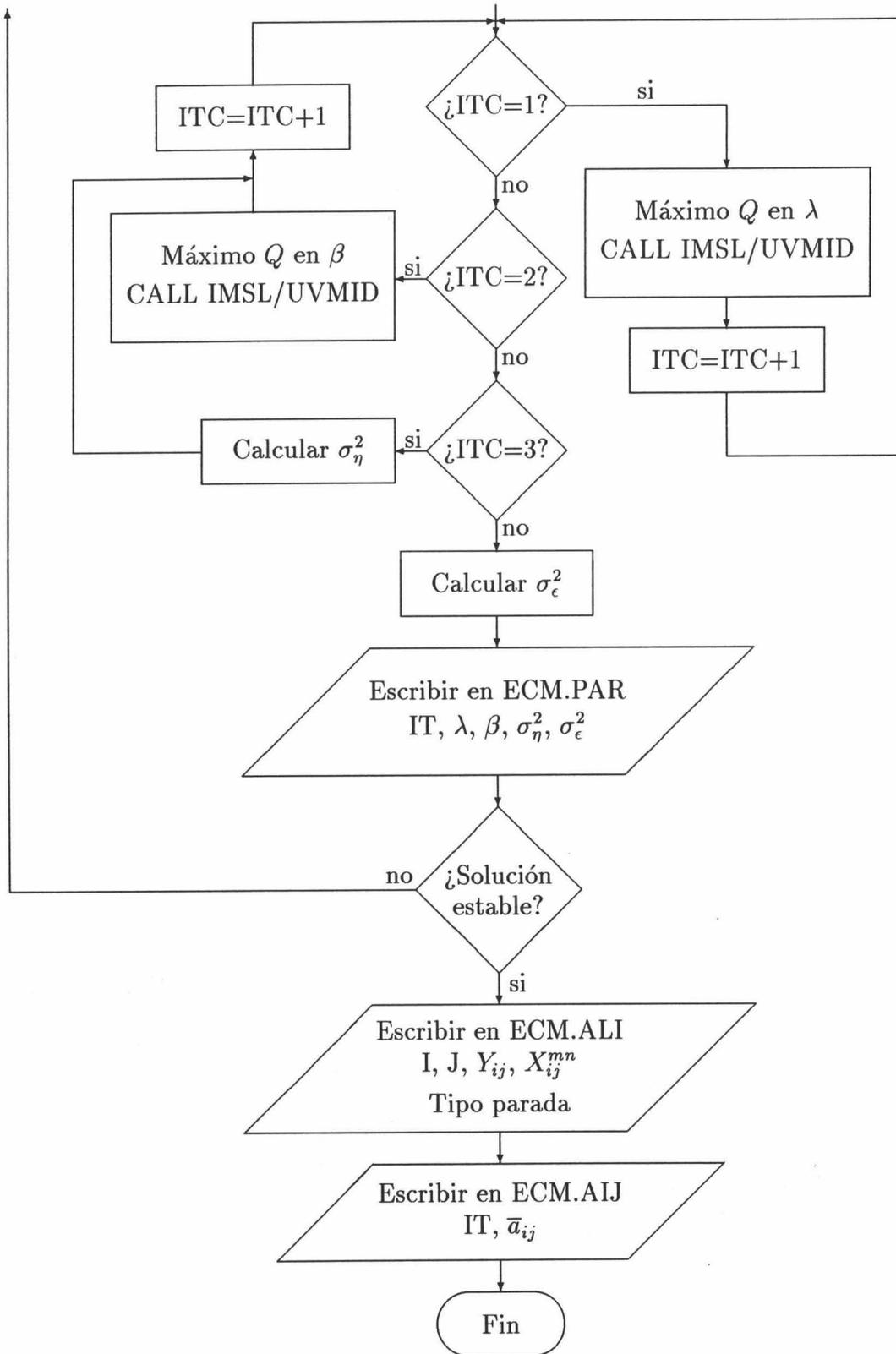


Apéndice C
Organigramas

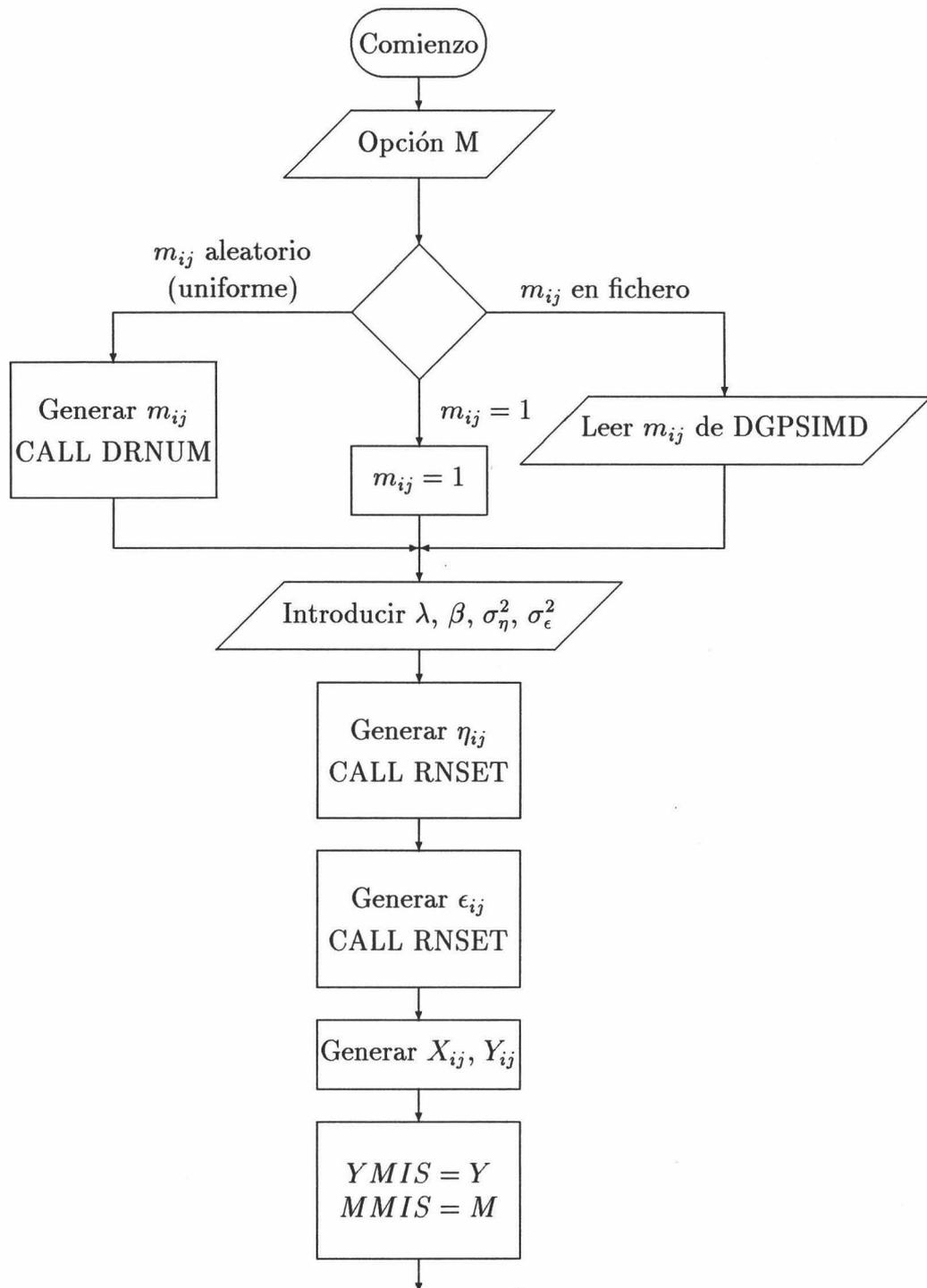


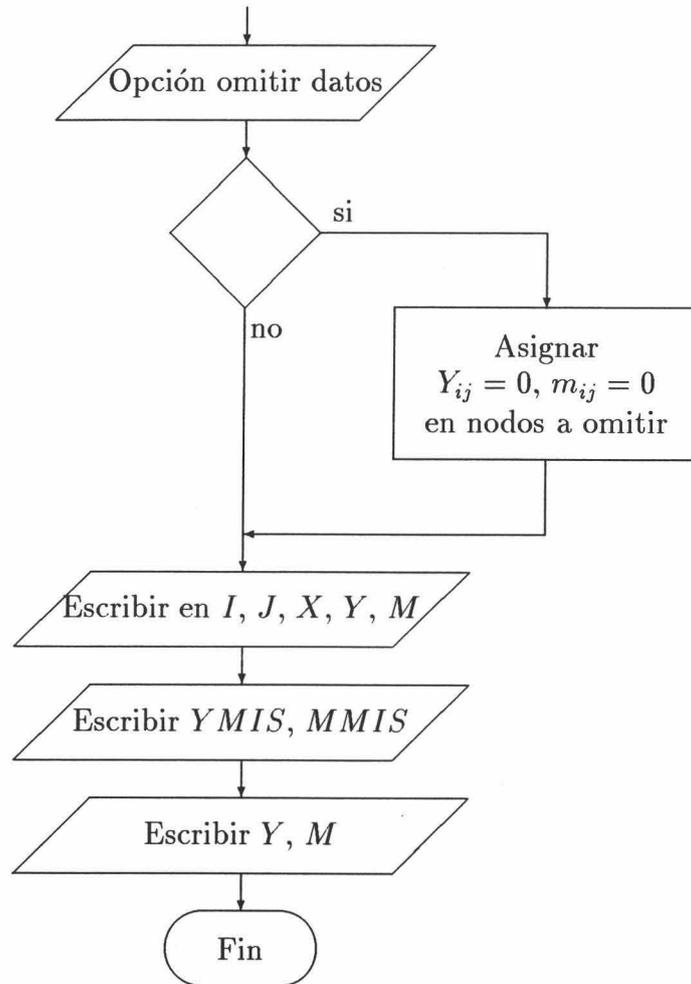
PROGRAMA ECMDGE





PROGRAMA SIMUDG







Bibliografía

- [1] Alonso, F.J., Angulo, J.M., y Shumway, R.H. (1993), "On numerical Fourier approximations for a certain type of spatial processes". *Applied Stochastic Models and Data Analysis*, Vol. I, pp. 17-22.
- [2] Alonso, F.J., Angulo, J.M., y Bueso, M.C. (1994), "Estimation and smoothing from incomplete data for a class of lattice processes". *Technical Report num.* Universidad de Granada, Dpto. Estadística e I.O..
- [3] Angulo, J.M., et al. (1992), "Discrete spatial Fourier approximations for parameter estimation and smoothing". Contributed paper to the *Seventh International Meeting on Multivariate Analysis* (University Park, Pennsylvania, May 1992) (unpublished).
- [4] Angulo, J.M., Bueso, M.C., y Alonso, F.J. (1993), "Estimation and smoothing for doubly geometric lattice processes with missing observations". *Contributed papers to the 49th session of ISI, Firenze Book 1*, pp. 37-38.
- [5] Angulo, J.M., Azari, A.S., Shumway, R.H. y Yucel, Z.T. (1994), "Fourier approximations for estimation and smoothing of irregularly observed spatial processes". *Proceedings of the International Conference on Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Kluwer Academic, por aparecer.
- [6] Bartlett, M.S. (1955), *An Introduction to Stochastic Processes*. Cambridge University Press, London.
- [7] Bartlett, M.S. (1975), *The Statistical Analysis of Spatial Pattern*. Chapman and Hall, London.

- [8] Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice system (with discussion)". *J. Roy. Stat. Soc., Ser. B.* 36, pp. 192–236.
- [9] Besag, J. (1977), "Errors-in-variables Estimation for Gaussian Lattice Schemes". *J. Roy. Stat. Soc., Ser. B.* 39, pp. 73–78.
- [10] Besag, J. (1981), "On a system of two-dimensional recurrence equations". *J. Roy. Stat. Soc., Ser. B.* 43, pp. 302–9.
- [11] Besag, J. (1986), "On the statistical analysis of dirty pictures". *J. Roy. Stat. Soc., Ser. B.* 48, pp. 259–279.
- [12] Besag, J. y Moran, P.A.P. (1975), "On the estimation and testing of spatial interaction in Gaussian lattice processes". *Biometrika* 62, pp. 555–62.
- [13] Boyles, R.A. (1983), "On the convergence of the EM algorithm". *J. Roy. Stat. Soc., Ser. B.* 45, pp. 47–50.
- [14] Brockwell, P.J. y Davis, R.A. (1991), *Time Series: Theory and Methods*. Springer-Verlag, New York.
- [15] Bueso, M.C., Alonso, F.J. y Angulo, J.M. (1994). "Estimación y alisamiento a partir de datos incompletos para una clase de procesos espaciales". *Actas del XXI congreso de la S.E.I.O.*, pp. 374–375.
- [16] Cairoli, R., Walsh, J.B. (1975). "Stochastic Integrals in the plane". *Acta Mathematica*, Vol. 134, 111–183.
- [17] Cliff, A.D. y Ord, J.K. (1981). *Spatial Processes: Models and Applications*. Pion, London.
- [18] Cressie, N.A.C. (1991), *Statistics for Spatial Data*. Wiley, New York.
- [19] Dempster, A.P., Laird, N.M. y Rubin, D.B. (1977), "Maximum likelihood from incomplete data via the EM algorithm". *J. Roy. Statist. Soc., Ser. B.* 39, pp. 1–38.
- [20] Gupta, N.K. y Mehra, R.K. (1974), "Computational Aspects of Maximun Likelihood Estimation and Reduction in Sensitivity Calculations". *IEEE Trans. Aut. Cont.* AC-19, pp. 911–924.

- [21] Guyon, X. (1982), "Parameter estimation for a stationary process on a d -dimensional lattice". *Biometrika* 69, 95–105.
- [22] Guyon, X. (1990), "Champs stationnaires sur \mathbb{Z}^2 : modeles, statistiques et simulations". *Contribuciones en probabilidad y estadística matemática* 1.
- [23] Guyon, X. (1992), *Champs aléatoires sur un réseau. Modélisations, statistique et applications*. Masson, Paris.
- [24] Haining, R.P. (1978), "The moving average models for spatial interaction". *Transactions, Institute of British Geographers N. S.* 3, pp. 205–225.
- [25] Heine, V. (1955), "Models for two-dimensional stationary stochastic processes". *Biometrika*, 42, pp. 170–178.
- [26] Huang, D. (1984), "The Autoregressive Moving Average Model for Spatial Analysis". *Austral. J. Statist.* 26(2), pp. 169–178.
- [27] Huang, D. y Anh, V.V. (1992), "Estimation of spatial ARMA models". *Austral. J. Statist.* 34(3), pp. 513–530.
- [28] Jazwinski, A.H. (1970), *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- [29] Jones, R.H. (1989), "Fitting a stochastic partial differential equation to aquifer data". *Stochastic Hydrologic and Hydraulics*, 3, pp. 85–96.
- [30] Kalman, R.E. (1960), "A new approach to linear filtering and prediction problems". *Trans. ASME J. of Basic Eng.*, 8, pp. 35–45.
- [31] Kalman, R.E. y Buzy, R.S. (1961), "New results in linear filtering and prediction theory". *Trans. ASME J. of Basic Eng.*, 83, pp. 95–108.
- [32] Katayama, T. (1979), "Restoration of noisy images using a two-dimensional linear model". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-9, N° 11, pp. 711–716.
- [33] Little, R.J.A. y Rubin, D.B. (1987), *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.

- [34] Mardia, K.V. (1988), "Multi-dimensional multivariate Gaussian Markov random fields with application to image processing". *Journal of Multivariate Analysis* 24, 265–284.
- [35] Mardia, K.V. y Marshall, R.J. (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression". *Biometrika* 71, 135–46.
- [36] Martin, R.J. (1979), "A subclass of lattice processes applied to a problem in planar sampling". *Biometrika*, 66, 2, pp. 209–217.
- [37] Martin, R.J. (1986), "On the design of experiments under spatial correlation". *Biometrika* 73, 247–277. Correction (1988) 75, 396.
- [38] Martin, R.J. (1990), "The use of time-series models and methods in the analysis of agricultural field trials". *Commun. Statist.—Theory Meth.*, 19 (1), pp. 55–81.
- [39] Meng, X.L. (1994), "On the rate of Convergence of the ECM algorithm". Por aparecer en *The Annals of Statistics* 22.
- [40] Meng, X.L. y Rubin, D.B. (1991). "Using EM to Obtain Asymptotic Variance–Covariance Matrices: The SEM Algorithm". *Journal of the American Statistical Association*, Vol. 86, No. 416, pp. 899–909.
- [41] Meng, X.L. y Rubin, D.B. (1993), "Maximum likelihood estimation via the ECM algorithm: A general framework". *Biometrika*, 80, pp. 267–278.
- [42] Moran, P.A.P. (1973a), "A Gaussian Markovian process on a square lattice". *Journal of Applied Probability* 10, pp. 54–62.
- [43] Moran, P.A.P. (1973b), "Necessary conditions for Markovian process on a lattice". *Journal of Applied Probability* 10, pp. 605–612.
- [44] Nualart, D. y Sanz, M. (1979), "A Markov property for two parameter Gaussian Processes". *Stochastica*, Vol. III num. 1, pp. 1–16.
- [45] Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction". *Journal of the American Statistical Association*, Vol. 70, No. 349, pp. 120–126.

- [46] Pawitan, Y., y Shumway, R.H. (1989), "Spectral estimation and deconvolution for a linear time series model". *Journal of Time Series Analysis*, 10(2), pp. 115–129.
- [47] Priestley, M.B.(1981), *Spectral Analysis and Time Series, Vol 1: Univariate Series*. Academic Press, New York.
- [48] Priestley, M.B.(1981), *Spectral Analysis and Time Series, Vol 2: Multivariate Series, Prediction and Control*. Academic Press, New York.
- [49] Ripley, B.D. (1981), *Spatial Statistic*. Wiley, New York.
- [50] Sanz, M. (1978), "Cálculo diferencial estocástico para procesos con parámetro n -dimensional". *Stochastica*, Vol. II num. 4, pp. 51–70.
- [51] Shumway, R.H. (1988), *Applied Statistical Time Series Analysis*. Englewood Cliffs, N J: Prentice–Hall.
- [52] Shumway, R. H., y Stoffer, D.S. (1981), "Time Series Smoothing and Forecasting Using using the EM algorithm". Technical Report num. 27, Division of Statistics, University of California, Davis.
- [53] Shumway, R. H., y Stoffer, D.S. (1982), "An approach to time series smoothing y forecasting using the EM algorithm". *J. Time Series Anal.* 3, pp. 253–264.
- [54] Stoffer, D.S. (1982), "Estimation of parameters in a linear dynamic system with missing observations", PH.D.Dissertation, University of California, Davis.
- [55] Tanner, M.A. (1993), *Tools for Statistical Inference. Observed data and data augmentation methods*. Springer–Verlag, New York.
- [56] Tjøstheim, D. (1978), "Statistical spatial series modelling". *Adv. Appl. Prob.*, 10, pp. 130–154.
- [57] Tjøstheim, D. (1983), "Statistical spatial series modelling II: Some further results on unilateral lattice processes", *Adv. Appl. Prob.*, 15, pp. 562–584.
- [58] Todorovic, P. (1992), *An Introduction to Stochastic Processes and Their Applications*. Springer-Verlag, New York.

- [59] Unny, T.E. (1989), "Stochastic partial differential equations in ground-water hydrology. Part I: Theory". *Stochastic Hydrology and Hydraulics* 3, pp. 135–153.
- [60] vanDik, D.A., Meng, X.L., y Rubin, D.B. (1994), "Maximun Likelihood Estimation via the ECM algorithm: Computing the Asymptotic Variance". *Technical Report num. 380*, Departament of Statistics, University of Chicago.
- [61] Vecchia, A.V. (1988), "Estimation and model identification for continuous spatial processes". *J. Roy. Statist. Soc., Ser. B.* 50, pp. 297–312.
- [62] Whittle, P. (1954), "On stationary processes in the plane". *Biometrika*, 41, pp. 434–449.
- [63] Whittle, P. (1963), *Prediction and Regulation by Linear Least-square Methods*. English Universities Press, London.
- [64] Woods, J.W., y C.H. Radewan (1977), "Kalman filtering in two dimensions". *IEEE Trans. Inform. Theory*, Vol. IT-23, pp. 473–482.
- [65] Wu, C.F.J. (1983), "On the convergence properties of the EM algorithm". *The Annals of Statistics*, Vol. 11, No. 1, pp. 95–103.
- [66] Yadrenko, M.I. (1983), *Spectral Theory of Random Fields*. Springer-Verlag, New York.